GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:39:49 ; Search time 1821.26 Seconds
(without alignments)
6181.697 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 2011660

Minimum DB seq length: 0
Maximum DB seq length: 538

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	538	100.0	538	6	AR028488	AR028488 Sequence
	2	538	100.0	538	6	I55851	I55851 Sequence 14
	3	530	98.5	530	9	HSU19143	U19143 Human GAGE-
	4	525.4	97.7	528	9	AF055473	AF055473 Homo sapi
	5	504.8	93.8	532	6	AR028491	AR028491 Sequence
	6	504.8	93.8	532	6	I55854	I55854 Sequence 17
	7	501.4	93.2	528	6	AX334151	AX334151 Sequence
	8	501.4	93.2	528	9	HSU19145	U19145 Human GAGE-
	9	498.8	92.7	527	9	HSU19147	U19147 Human GAGE-
	10	497.6	92.5	524	9	AF058988	AF058988 Homo sapi
	11	496.8	92.3	524	9	HSU19146	U19146 Human GAGE-
	12	488	90.7	526	9	AF055474	AF055474 Homo sapi
c	13	301.4	56.0	530	6	AX284300	AX284300 Sequence
c	14	299.2	55.6	365	6	AX284692	AX284692 Sequence
	15	182.4	33.9	493	9	HSA318881	AJ318881 Homo sapi
	16	130.4	24.2	475	6	AX226497	AX226497 Sequence
	17	118.6	22.0	448	6	AX226538	AX226538 Sequence
	18	111.4	20.7	529	9	BC009538	BC009538 Homo sapi
	19	109	20.3	448	6	AX226537	AX226537 Sequence
	20	108	20.1	399	6	AX062442	AX062442 Sequence
	21	104.4	19.4	463	9	HSA290447	AJ290447 Homo sapi
	22	102	19.0	493	9	BC010897	BC010897 Homo sapi
	23	101.6	18.9	515	6	AX106691	AX106691 Sequence
	24	101.6	18.9	515	6	AX140982	AX140982 Sequence
	25	101.6	18.9	515	6	AX200842	AX200842 Sequence
	26	101.6	18.9	515	6	AX267498	AX267498 Sequence
	27	100.8	18.7	214	6	AX208978	AX208978 Sequence
	28	100.8	18.7	320	6	AX198429	AX198429 Sequence
	29	98.4	18.3	491	9	AF275258	AF275258 Homo sapi
c	30	93.4	17.4	346	6	AX337333	AX337333 Sequence
	31	93.4	17.4	476	9	HSA005894	AJ005894 Homo sapi
	32	92.8	17.2	457	6	AX063163	AX063163 Sequence
c	33	86.4	16.1	348	6	AX226603	AX226603 Sequence
	34	85.6	15.9	481	9	HSA318879	AJ318879 Homo sapi
	35	84.8	15.8	479	6	AX062443	AX062443 Sequence
	36	82.8	15.4	458	6	AX226495	AX226495 Sequence
	37	77.4	14.4	461	6	AX063165	AX063165 Sequence
c	38	53.6	10.0	445	11	G41485	G41485 Z4192 Zebra
	39	51	9.5	51	6	AX204315	AX204315 Sequence
	40	49.4	9.2	51	6	AX204507	AX204507 Sequence
	41	49	9.1	49	6	AX226542	AX226542 Sequence
	42	48.8	9.1	110	6	AX226525	AX226525 Sequence
	43	48.8	9.1	122	6	AX226530	AX226530 Sequence
	44	45.8	8.5	209	9	HSA318895	AJ318895 Homo sapi
	45	45.6	8.5	110	6	AX226526	AX226526 Sequence

ALIGNMENTS

RESULT 1
AR028488
LOCUS AR028488 538 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 14 from patent US 5858689.
ACCESSION AR028488
VERSION AR028488.1 GI:5940461
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 538)
AUTHORS van der Bruggen,P., van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
TITLE Isolated peptides derived from the gage tumor rejection antigen precursor and uses thereof
JOURNAL Patent: US 5858689-A 14 12-JAN-1999;
FEATURES Location/Qualifiers
source 1. .538
/organism="unknown"
BASE COUNT 160 a 116 c 155 g 107 t
ORIGIN

Query Match 100.0%; Score 538; DB 6; Length 538;
Best Local Similarity 100.0%; Pred. No. 6.6e-120;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
```

RESULT 2
I55851
LOCUS I55851 538 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 14 from patent US 5648226.
ACCESSION I55851
VERSION I55851.1 GI:2476645
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 538)
AUTHORS Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
TITLE Isolated peptides derived from tumor rejection antigens, and their use
JOURNAL Patent: US 5648226-A 14 15-JUL-1997;
FEATURES Location/Qualifiers
source 1. .538
/organism="unknown"
BASE COUNT 160 a 116 c 155 g 107 t
ORIGIN

Query Match 100.0%; Score 538; DB 6; Length 538;
Best Local Similarity 100.0%; Pred. No. 6.6e-120;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


```
Db      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
```


RESULT 3
HSU19143
LOCUS HSU19143 530 bp mRNA linear PRI 04-DEC-1995
DEFINITION Human GAGE-2 protein mRNA, complete cds.
ACCESSION U19143
VERSION U19143.1 GI:914900
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 530)
AUTHORS Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S. and Boon,T.
TITLE A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma
JOURNAL J. Exp. Med. 182 (3), 689-698 (1995)
MEDLINE 95378788
REFERENCE 2 (bases 1 to 530)
AUTHORS Van Den Eynde,B.J.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1994) Benoit J Van Den Eynde, Ludwig Institute For Cancer Research, 74 Avenue Hippocrate, BRUSSELS, 1200, BELGIUM
FEATURES Location/Qualifiers
source 1. .530
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="female"
/cell_line="MZ2-MEL.43"
/tissue_type="melanoma"
/dev_stage="adult"
CDS 84. .434
/codon_start=1
/product="GAGE-2 protein"
/protein_id="AAA82745.1"
/db_xref="GI:914901"
/translation="MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGE
PATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPE
EVKTPEEGEKQSQC"
BASE COUNT 152 a 116 c 155 g 107 t
ORIGIN

```
  Query Match             98.5%;  Score 530;  DB 9;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-118;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530


RESULT   4
AF055473
LOCUS       AF055473                 528 bp    mRNA    linear   PRI 01-MAY-2000
DEFINITION  Homo sapiens GAGE-8 mRNA, complete cds.
ACCESSION   AF055473
VERSION     AF055473.1  GI:3511022
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   De Backer,O., Arden,K.C., Boretti,M., Vantomme,V., De Smet,C.,
            Czekay,S., Viars,C.S., De Plaen,E., Brasseur,F., Chomez,P., Van den
            Eynde,B., Boon,T. and van der Bruggen,P.
  TITLE     Characterization of the GAGE genes that are expressed in various
            human cancers and in normal testis
  JOURNAL   Cancer Res. 59 (13), 3157-3165 (1999)
  MEDLINE   99323388
   PUBMED   10397259
REFERENCE   2  (bases 1 to 528)
  AUTHORS   De Backer,O.R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1998) Ludwig Institute for Cancer Research,
            Brussels Branch, 74 av. Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             73. .423
                     /codon_start=1
                     /product="GAGE-8"
                     /protein_id="AAC33676.1"
                     /db_xref="GI:3511023"
                     /translation="MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGE
                     PATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPE
                     EVKTPEEGEKQSQC"
BASE COUNT      158 a    112 c    151 g    107 t
ORIGIN


  Query Match             97.7%;  Score 525.4;  DB 9;  Length 528;
  Best Local Similarity   99.8%;  Pred. No. 7.3e-117;
  Matches  526;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      12 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 71
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 60

Qy      72 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTAC 131
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTAC 120

Qy     132 GTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAA 191
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAA 180

Qy     192 CCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAG 251
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAG 240

Qy     252 GAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAG 311
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     241 GAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAG 300

Qy     312 GAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATG 371
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATG 360

Qy     372 GACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGT 431
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGT 420

Qy     432 TAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAA 491
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAA 480

Qy     492 ATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 527


RESULT   5
AR028491
LOCUS       AR028491                 532 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 17 from patent US 5858689.
ACCESSION   AR028491
VERSION     AR028491.1  GI:5940464
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   van der Bruggen,P., van den Eynde,B., DeBacker,O. and
            Boon-Falleur,T.
  TITLE     Isolated peptides derived from the gage tumor rejection antigen
            precursor and uses thereof
  JOURNAL   Patent: US 5858689-A 17 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="unknown"
BASE COUNT      156 a    111 c    154 g    111 t
ORIGIN
```

```
  Query Match             93.8%;  Score 504.8;  DB 6;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 7.1e-112;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| .
Db       1 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           |||||||||||||||||||||||||||||||||||||   ||| ||||||||||||| |
Db      61 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGC 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     121 GCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 180

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 240

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     241 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATA 300

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 360

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 420

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 480

Qy     487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 532
```

```
RESULT   6
I55854
LOCUS       I55854                   532 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 17 from patent US 5648226.
ACCESSION   I55854
VERSION     I55854.1  GI:2476648
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
  TITLE     Isolated peptides derived from tumor rejection antigens, and their
            use
  JOURNAL   Patent: US 5648226-A 17 15-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="unknown"
BASE COUNT      156 a    111 c    154 g    111 t
ORIGIN
```

```
  Query Match             93.8%;  Score 504.8;  DB 6;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 7.1e-112;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 120

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
```

```
           ||||||||||||||||||||||||||||||||||||||   ||| ||||||||||||| |
Db     121 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGC 180

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     181 GCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 240

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATA 300

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     301 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 360

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 420

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 480

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 532
```

```
RESULT   7
AX334151
LOCUS       AX334151                 528 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 4660 from Patent WO0194629.
ACCESSION   AX334151
VERSION     AX334151.1  GI:18124870
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 4660 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a    114 c    156 g    111 t
ORIGIN
```

```
  Query Match             93.2%;  Score 501.4;  DB 6;  Length 528;
  Best Local Similarity   98.3%;  Pred. No. 4.7e-111;
  Matches  518;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy       2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy      62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db      61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy     119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
           |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180
```

```
Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy    479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 527
```

```
RESULT   8
HSU19145
LOCUS       HSU19145                 528 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-4 protein mRNA, complete cds.
ACCESSION   U19145
VERSION     U19145.1  GI:914904
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 528)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1994) BENOIT J VAN DEN EYNDE, Ludwig Institute
            for Cancer Research, 74 Avenue Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             83. .436
                     /codon_start=1
                     /product="GAGE-4 protein"
                     /protein_id="AAA82747.1"
                     /db_xref="GI:914905"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      147 a    114 c    156 g    111 t
ORIGIN
```

```
  Query Match             93.2%;  Score 501.4;  DB 9;  Length 528;
  Best Local Similarity   98.3%;  Pred. No. 4.7e-111;
  Matches  518;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;
```

```
Qy      2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy     62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
          ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db     61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
          |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy    479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 527
```

```
RESULT   9
HSU19147
LOCUS       HSU19147                 527 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-6 protein mRNA, complete cds.
ACCESSION   U19147
VERSION     U19147.1  GI:914908
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 527)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1994) BENOIT J VAN DEN EYNDE, Ludwig Institute
            for Cancer Research, 74 Avenue Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             82. .435
                     /codon_start=1
                     /product="GAGE-6 protein"
                     /protein_id="AAA82749.1"
                     /db_xref="GI:914909"
```

```
                     /translation="MSWRGRSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEVDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      146 a    113 c    157 g    111 t
ORIGIN


  Query Match             92.7%;  Score 498.8;  DB 9;  Length 527;
  Best Local Similarity   98.1%;  Pred. No. 2e-110;
  Matches  516;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy       3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 60

Qy      63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||
Db      61 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGA 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||| ||||| ||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db     121 CCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGGCAGGAGGTGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
           |||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 526


RESULT  10
AF058988
LOCUS       AF058988                524 bp    mRNA    linear   PRI 11-JUL-1998
DEFINITION  Homo sapiens melanoma antigen related GAGE-7 mRNA, complete cds.
ACCESSION   AF058988
VERSION     AF058988.1  GI:3300089
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   Chen,M.E., Lin,S.-H., Chung,L.W.K. and Sikes,R.A.
  TITLE     Isolation and characterization of PAGE-1 and GAGE-7: new genes
            expressed in the LNCaP prostate cancer progression model that share
            homology with melanoma associated antigens
  JOURNAL   J. Biol. Chem. (1998) In press
REFERENCE   2  (bases 1 to 524)
  AUTHORS   Chen,M.E., Lin,S.-H., Chung,L.W.K. and Sikes,R.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-1998) Urology, University of Virginia, Box 422,
            Charlottesville, VA 22908, USA
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="LNCaP; C4-2"
                     /note="isolated from prostate cancer cell lines by
                     differential display PCR; expression pattern correlates
                     with progression from androgen sensitive to androgen
                     insensitive in the human prostate cancer progression model
                     LNCaP to C4-2"
     CDS             80. .433
                     /note="similar to GAGE family of melanoma antigens"
                     /codon_start=1
                     /product="melanoma antigen related GAGE-7"
                     /protein_id="AAC25989.1"
                     /db_xref="GI:3300090"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      146 a    113 c    152 g    111 t      2 others
ORIGIN


  Query Match             92.5%;  Score 497.6;  DB 9;  Length 524;
  Best Local Similarity   97.9%;  Pred. No. 3.9e-110;
  Matches  513;  Conservative    2;  Mismatches    6;  Indels    3;  Gaps    1;

Qy       5 CAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTG 64
           ||||||||||||||||||:|||||:||||||||||||||||||||||||||||||||||||
Db       1 CAGGGAGCTGTGAGGCAGTSCTGTGYGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTG 60

Qy      65 AGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACC 121
           ||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||||
Db      61 AGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACC 120

Qy     122 AAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 181
           ||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 180

Qy     182 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 240

Qy     242 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 300

Qy     302 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 360

Qy     362 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 420

Qy     422 ATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 481
           |||||||||||||||||| |||||||||||||||||||||||||||| ||||||||||
Db     421 ATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATTTTGGAAATTT 480

Qy     482 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
           |||||||||||||||||||||||||||||||||||||||||||||
Db     481 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 524


RESULT  11
HSU19146
LOCUS       HSU19146                524 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-5 protein mRNA, complete cds.
ACCESSION   U19146
VERSION     U19146.1  GI:914906
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
```

```
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 524)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1994) BENOIT J VAN DEN EYNDE, Ludwig Institute
            for Cancer Research, 74 Avenue Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             75. .428
                     /codon_start=1
                     /product="GAGE-5 protein"
                     /protein_id="AAA82748.1"
                     /db_xref="GI:914907"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      148 a    111 c    154 g    111 t
ORIGIN
```

```
  Query Match             92.3%;  Score 496.8;  DB 9;  Length 524;
  Best Local Similarity   98.1%;  Pred. No. 6.1e-110;
  Matches  514;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           ||||||||||||||||||||||||||||||||||||    |||  ||||||||||||| |
Db      61 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGC 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     121 GCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 180

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 240

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     241 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATA 300

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 360

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 420

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 480

Qy     487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530
           |||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 524

RESULT  12
```

```
AF055474
LOCUS       AF055474                526 bp    mRNA    linear   PRI 01-MAY-2000
DEFINITION  Homo sapiens GAGE-7B mRNA, complete cds.
ACCESSION   AF055474
VERSION     AF055474.1  GI:3511024
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 526)
  AUTHORS   De Backer,O., Arden,K.C., Boretti,M., Vantomme,V., De Smet,C.,
            Czekay,S., Viars,C.S., De Plaen,E., Brasseur,F., Chomez,P., Van den
            Eynde,B., Boon,T. and van der Bruggen,P.
  TITLE     Characterization of the GAGE genes that are expressed in various
            human cancers and in normal testis
  JOURNAL   Cancer Res. 59 (13), 3157-3165 (1999)
  MEDLINE   99323388
   PUBMED   10397259
REFERENCE   2  (bases 1 to 526)
  AUTHORS   De Backer,O.R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1998) Ludwig Institute for Cancer Research,
            Brussels Branch, 74 av. Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .526
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             55. .408
                     /codon_start=1
                     /product="GAGE-7B"
                     /protein_id="AAC33677.1"
                     /db_xref="GI:3511025"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      168 a    109 c    143 g    106 t
ORIGIN
```

```
  Query Match             90.7%;  Score 488;  DB 9;  Length 526;
  Best Local Similarity   98.4%;  Pred. No. 8.2e-108;
  Matches  504;  Conservative    0;  Mismatches    5;  Indels    3;  Gaps    1;

Qy      30 TGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGT 89
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGT 60

Qy      90 TGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAA 146
           ||||||||||||||||||   ||| |||||||||||||| ||||| ||| ||||||||||
Db      61 TGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAA 120

Qy     147 ATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAA 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAA 180

Qy     207 GAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAG 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAG 240

Qy     267 GGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCA 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCA 300

Qy     327 CAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCA 386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCA 360

Qy     387 GAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGT 446
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     361 GAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGT 420

Qy     447 TGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAG 506
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAG 480

Qy    507 CTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||
Db    481 CTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 512


RESULT 13
AX284300/c
LOCUS       AX284300                530 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 105 from Patent WO0179556.
ACCESSION   AX284300
VERSION     AX284300.1  GI:17044988
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
  TITLE     Novel genes, compositions and methods for the identification,
            assessment, prevention, and therapy of human cancers
  JOURNAL   Patent: WO 0179556-A 105 25-OCT-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       89 a    100 c    106 g    183 t     52 others
ORIGIN


  Query Match             56.0%;  Score 301.4;  DB 6;  Length 530;
  Best Local Similarity   87.9%;  Pred. No. 1.2e-62;
  Matches  320;  Conservative    0;  Mismatches   42;  Indels    2;  Gaps    1;

Qy    177 AGTGATGAAGTGGAACCAGC--AACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGG 234
          ||| || ||||||| ||| |  |   |||||| ||||| ||||||||||||||||| |||
Db    449 AGTNATNAAGTGGACCCANCCNACNCCCTGAANAAGGGNAACCAGCAACTCAACGTNAGG 390

Qy    235 ATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGC 294
          |||||| |||||| ||| ||||| ||||| |||| |||| ||||||||||||||||||||
Db    389 ATCCTGNAGCTGCNCAGNAGGGANAGGATNAGGGNGCATNTGCAGGTCAAGGGCCCAAGC 330

Qy    295 CTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTC 354
          |||||| | |||||||||||||||||| ||| || |||||| || |||| || ||||||
Db    329 CTGAAGNTGATAGCCAGGAACAGGGTCNCCCCCANACTGGGNGTNAGTGTNAAAATGGTC 270

Qy    355 CTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTG 414
          || |||||||| | |||||||| ||||||||| || |||||||||| |||||| ||||||
Db    269 CTNATGGGCAGNANATGGACCCNCCAAATCCAAAGNAGGTGAAAACNCCTGAANAAGGTG 210

Qy    415 AAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTG 474
          |||||||||| |||||||||| ||| | ||||||||||||||||||||||||||||||||
Db    209 AAAAGCAATCCCAGTGTTAAAANAAGGCNCGTTGAAATGATGCAGGCTGCTCCTATGTTG 150

Qy    475 GAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAA 534
          |||||||||||||||||||||||||||||||||||||||||| |  ||| ||||||||
Db    149 GAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTNTNAAAAAAAAAAAAA 90

Qy    535 AAAA 538
          ||||
Db     89 AAAA 86


RESULT 14
AX284692/c
LOCUS       AX284692                365 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 497 from Patent WO0179556.
ACCESSION   AX284692
VERSION     AX284692.1  GI:17045380
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
  TITLE     Novel genes, compositions and methods for the identification,
            assessment, prevention, and therapy of human cancers
  JOURNAL   Patent: WO 0179556-A 497 25-OCT-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .365
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       63 a     86 c     66 g    150 t
ORIGIN


  Query Match             55.6%;  Score 299.2;  DB 6;  Length 365;
  Best Local Similarity   95.9%;  Pred. No. 4e-62;
  Matches  307;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy    219 GCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCA 278
          ||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||||
Db    365 GCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAAAGGATAAGGGAGCATCTGCA 306

Qy    279 GGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGT 338
          |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db    305 GGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAGGAACAGGGTCACCCCCAGACTGGGTGT 246

Qy    339 GAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAA 398
          ||||||||| |||||||||||||||| |||||||||||||||||| ||||||||||
Db    245 GAGTGTGAAAATGGTCCTGATGGGCAGGAAATGGACCCGCCAAATCCAAAGGAGGTGAAA 186

Qy    399 ACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCA 458
          |||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||||
Db    185 ACGCCTGAAGAAGGTGAAAAGCAATCTCAGTGTTAAAAAAAGGCACGTTGAAATGATGCA 126

Qy    459 GGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTT 518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 GGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTT 66

Qy    519 CTGCAAAGAAAAAAAAAAAA 538
          ||  ||| ||||||||||||
Db     65 CTAAAAAAAAAAAAAAAAAA 46


RESULT 15
HSA318881
LOCUS       HSA318881               493 bp    mRNA    linear   PRI 14-JAN-2002
DEFINITION  Homo sapiens mRNA for XAGE-3 protein.
ACCESSION   AJ318881
VERSION     AJ318881.1  GI:18157211
KEYWORDS    XAGE-3 gene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Zendman,A.J.W., van Kraats,A.A., Weidle,U.H., Ruiter,D.R. and Van
            Muijen,G.N.P.
  TITLE     Expression profile of members of the XAGE cancer/testis antigen
            family
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 493)
  AUTHORS   Zendman,A.J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-2001) Zendman A.J.W., Department of Pathology,
            University Medical Center St Radboud, Geert Grooteplein Zuid 24, PO
            Box 9101 6500 HB Nijmegen, NETHERLANDS
FEATURES             Location/Qualifiers
     source          1. .493
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp11.21-Xp11.22"
                     /tissue_type="Ewing sarcoma"
     gene            61. .396
                     /gene="XAGE-3"
     CDS             61. .396
                     /gene="XAGE-3"
                     /codon_start=1
                     /product="XAGE-3 protein"
                     /protein_id="CAC83008.1"
                     /db_xref="GI:18157212"
                     /translation="MIWRGRSTYRPRPRRSVPPPELIGPMLEPGDEEPQQEEPPTESR
                     DPAPGQEREEDQGAAETQVPDLEADLQELSQSKTGGECGNGPDDQGKILPKSEQFKMP
                     EGGDRQPQV"
BASE COUNT      167 a     98 c    125 g    103 t
ORIGIN


  Query Match             33.9%;  Score 182.4;  DB 9;  Length 493;
  Best Local Similarity   68.9%;  Pred. No. 7.3e-34;
  Matches  317;  Conservative    0;  Mismatches  121;  Indels   22;  Gaps    4;

Qy      76 TGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTACGTAG 135
           |||||||||||| |||||||||||||| || ||| ||||||| || ||  |    |||
Db      53 TGTGAAATATGATTTGGCGAGGAAGATCAACATATAGGCCTAGGCCGAGGAGAAGTGTAC 112

Qy     136 AGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAG 195
           ||||||||  ||||||||||||||||        | ||  | ||||||| | |
Db     113 CACCTCCTGAGCTGATTGGGCCTATGC------TGGAGCCCGGTGATGAGGAG------- 159

Qy     196 CAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAG- 254
               ||| |  ||| ||||||| ||||| || ||| ||||||||| ||| |||||||
Db     160 -----CCTCAGCAAGAGGAACCACCAACTGAAAGTCGGGATCCTGCACCTGGTCAGGAGA 214

Qy     255 --GGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGG 312
           | ||| ||| |||| ||| |||    ||||| |||  | |  |||||| ||  |||||
Db     215 GAGAAGAAGATCAGGGTGCAGCTGAGACTCAAGTGCCTGACCTGGAAGCTGATCTCCAGG 274

Qy     313 AACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGG 372
           | | |  |||  || |||||||| |||| |||| | |||||||||||  |||| || |
Db     275 AGCTGTCTCAGTCAAAGACTGGGGGTGAATGTGGAAATGGTCCTGATGACCAGGGGAAGA 334

Qy     373 ACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTT 432
           | ||||||  ||||  |  | |||| ||| |||| |||||| | |||| |||||  ||
Db     335 TTCTGCCAAAATCAGAACAATTTAAAATGCCAGAAGGAGGTGACAGGCAACCACAGGTTT 394

Qy     433 AAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAA 492
           ||| ||||||| | |||||  |  ||      |  | | |   | || | |   ||||||
Db     395 AAATGAAGACAAGCTGAAACAACACAAAACTGT-TTTTATCTAAGATATTTGACTTAAAA 453

Qy     493 TTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAA 532
           | ||  ||||||  ||| |||| |||| |||| ||||||
Db     454 ATATCGAAATAAACTTTTGCAGCTTTCTCCAAAAAAAAAA 493


Search completed: July  1, 2002, 07:41:59
Job time: 3730 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:37:34 ; Search time 1595.88 Seconds
(without alignments)
4550.068 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 17277834

Minimum DB seq length: 0
Maximum DB seq length: 538

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	492	91.4	527	10	BI826605	BI826605 603077056
	2	476.8	88.6	509	10	BI868671	BI868671 603392594
c	3	452.4	84.1	489	9	AA447559	AA447559 zw81e11.s
c	4	431.4	80.2	457	9	AW510753	AW510753 hd39d05.x
c	5	424.6	78.9	447	9	AI381509	AI381509 te76b07.x
c	6	424.4	78.9	464	9	AA738037	AA738037 nx15e11.s
c	7	413.8	76.9	418	9	AW016546	AW016546 UI-H-BI0p
	8	413.4	76.8	480	10	BG120336	BG120336 602353732
c	9	412.8	76.7	426	9	AA868226	AA868226 ak48h07.s
c	10	410.8	76.4	450	9	AA760996	AA760996 nx32h08.s
c	11	401.2	74.6	455	9	AI187350	AI187350 qf29a05.x
c	12	365.4	67.9	412	9	AW102587	AW102587 xd67g07.x
	13	360.4	67.0	445	11	BC005363	BC005363 Homo sapi
	14	359.4	66.8	419	9	AA448542	AA448542 zw81e11.r
c	15	329.8	61.3	384	9	AA913206	AA913206 om53f03.s
	16	318.8	59.3	397	10	BG206349	BG206349 RST25797
c	17	296.4	55.1	398	9	AA918604	AA918604 ol53f02.s
c	18	250.4	46.5	256	10	BG181480	BG181480 RST328 At
	19	233.4	43.4	258	10	BG186708	BG186708 RST5683 A
	20	228.2	42.4	253	10	BG184057	BG184057 RST2973 A
	21	226.6	42.1	261	10	BG199060	BG199060 RST18338
	22	224.8	41.8	245	10	BG212621	BG212621 RST32216
	23	219.8	40.9	292	10	BG212622	BG212622 RST32217
	24	214.8	39.9	245	10	BG208433	BG208433 RST27934
	25	213.4	39.7	524	11	AF318372	AF318372 Homo sapi
c	26	207.6	38.6	507	9	AA972716	AA972716 op90e12.s
	27	207	38.5	256	10	BG216461	BG216461 RST36040
c	28	204.8	38.1	224	9	AA738394	AA738394 nx28c12.s
c	29	202.4	37.6	382	10	BF869799	BF869799 IL3-ET011
	30	200.6	37.3	499	10	BG354572	BG354572 PLAC6 Hum
c	31	193.8	36.0	197	9	AI968311	AI968311 wu01f07.x
c	32	192.8	35.8	226	10	BG220441	BG220441 RST40225
c	33	186.8	34.7	503	9	AI742551	AI742551 wg55c11.x
c	34	183.8	34.2	468	9	AI128458	AI128458 qc39a12.x
c	35	181.8	33.8	453	9	AI189715	AI189715 qd28e08.x
c	36	178.4	33.2	483	9	AI149003	AI149003 qc75a02.x
c	37	177.8	33.0	496	9	AI188535	AI188535 qd09e11.x
c	38	176.8	32.9	468	9	AA039331	AA039331 zk39e06.s
c	39	174.4	32.4	496	9	AA993199	AA993199 ot93c08.s
c	40	173.6	32.3	280	10	X93834	X93834 HSIMBB135 H
c	41	172.4	32.0	445	10	BE302172	BE302172 bb82d02.x
c	42	171	31.8	532	9	AA737311	AA737311 nw51g04.s
c	43	170.4	31.7	429	9	AI189067	AI189067 qd18a09.x
c	44	170	31.6	437	9	AI740470	AI740470 wg15g08.x
	45	169.6	31.5	394	10	D59091	D59091 HUM521C10B

ALIGNMENTS

RESULT 1
BI826605
LOCUS BI826605 527 bp mRNA linear EST 04-OCT-2001
DEFINITION 603077056F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5168892 5', mRNA sequence.
ACCESSION BI826605
VERSION BI826605.1 GI:15938155
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 527)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11419 row: p column: 13
High quality sequence stop: 519.
FEATURES Location/Qualifiers
source 1. .527
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5168892"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range

0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH_MGC Library."
BASE COUNT 154 a 111 c 151 g 111 t
ORIGIN

Query Match 91.4%; Score 492; DB 10; Length 527;
Best Local Similarity 98.3%; Pred. No. 3.8e-104;
Matches 519; Conservative 0; Mismatches 5; Indels 4; Gaps 2;

```
Qy     12 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 71
Db      1 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 60

Qy     72 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGC 128
Db     61 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGC 120

Qy    129 TACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTG 188
Db    121 TATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTG 180

Qy    189 GAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCT 248
Db    181 GAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCT 240

Qy    249 CAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGC 308
Db    241 CAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGC 300

Qy    309 CAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAG 368
Db    301 CAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAG 360

Qy    369 ATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAG 428
Db    361 ATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAG 420

Qy    429 TGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATT 488
Db    421 TGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATT 480

Qy    489 AAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAA 536
Db    481 AAAATTCTCCCAATAAAGCTTTACAG-CTTCTGCAAAGAAAAAAAAA 527
```

RESULT 2
BI868671
LOCUS BI868671 509 bp mRNA linear EST 11-OCT-2001
DEFINITION 603392594F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5402663 5', mRNA sequence.
ACCESSION BI868671
VERSION BI868671.1 GI:16042344
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 509)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
• cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12026 row: l column: 24
High quality sequence stop: 509.
FEATURES Location/Qualifiers
source 1. .509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5402663"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 142 a 107 c 154 g 106 t
ORIGIN

Query Match 88.6%; Score 476.8; DB 10; Length 509;
Best Local Similarity 98.0%; Pred. No. 1.3e-100;
Matches 494; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

```
Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
Db      6 ACGCTAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 65

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
Db     66 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTA 125

Qy    118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
Db    126 GACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 185

Qy    178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
Db    186 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 245

Qy    238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
Db    246 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCTGAAGCCTG 305

Qy    298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
Db    306 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 365

Qy    358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 417
Db    366 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 425

Qy    418 AGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 477
Db    426 AGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 485

Qy    478 ATTTGTTCATTAAAATTCTCCCAA 501
Db    486 ATTTGTTCATTAAAATTCTCCCAA 509
```

RESULT 3
AA447559/c
LOCUS AA447559 489 bp mRNA linear EST 04-JUN-1997
DEFINITION zw81e11.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782636 3' similar to TR:G914905 G914905 GAGE-4 PROTEIN. [1] ;, mRNA sequence.
ACCESSION AA447559
VERSION AA447559.1 GI:2161229
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 489)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:782636"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      101 a    139 c    105 g    144 t
ORIGIN
```

```
  Query Match             84.1%;  Score 452.4;  DB 9;  Length 489;
  Best Local Similarity   98.0%;  Pred. No. 6e-95;
  Matches  480;  Conservative    0;  Mismatches    6;  Indels    4;  Gaps    2;

Qy      43 CCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGAT 102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 CCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGAT 430

Qy     103 CGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTA 159
           |||||   ||| ||||||||||||| ||||| ||| |||||||||||| ||||| ||||||
Db     429 CGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAAGTGATT-GGCCTA 371

Qy     160 TGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     370 TGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 311

Qy     220 CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 251

Qy     280 GTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     250 GTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 191

Qy     340 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 131

Qy     400 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAG 459
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     130 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAG 71

Qy     460 GCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTC 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 GCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTC 11

Qy     520 TGCAAAGAAA 529
           ||||||||||
Db      10 TGCAAAGAAA 1
```

```
RESULT   4
AW510753/c
LOCUS       AW510753                 457 bp    mRNA    linear   EST 03-MAR-2000
DEFINITION  hd39d05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2911881 3' similar to SW:GGE4_HUMAN Q13068 GAGE-4 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AW510753
VERSION     AW510753.1  GI:7148831
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2911881"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       87 a    133 c     95 g    142 t
ORIGIN
```

```
  Query Match             80.2%;  Score 431.4;  DB 9;  Length 457;
  Best Local Similarity   98.0%;  Pred. No. 4.5e-90;
  Matches  448;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      77 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATC---GGCCTAGACCAAGACGCTACGT 133
           ||||||||||||||||||||||||||||||||    |||||||||||| ||||| ||
Db     457 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGT 398

Qy     134 AGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACC 193
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     397 ACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACC 338

Qy     194 AGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGA 253
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     337 AGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGA 278

Qy     254 GGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGA 313
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     277 GGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGA 218
```

```
Qy     314 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGA 373
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     217 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGA 158

Qy     374 CCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTA 433
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     157 CCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTA 98

Qy     434 AAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAAT 493
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      97 AAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAAT 38

Qy     494 TCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530
           |||||||||||||||||||||||||||||| ||||
Db      37 TCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAA 1
```

```
RESULT   5
AI381509/c
LOCUS       AI381509                447 bp    mRNA    linear   EST 28-MAR-1999
DEFINITION  te76b07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2092597 3' similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AI381509
VERSION     AI381509.1  GI:4194290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 497   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2092597"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       84 a    130 c     95 g    138 t
ORIGIN
```

```
  Query Match           78.9%;  Score 424.6;  DB 9;  Length 447;
  Best Local Similarity  98.4%;  Pred. No. 1.7e-88;
  Matches  440;  Conservative    0;  Mismatches    4;  Indels    3;  Gaps    1;

Qy      88 GTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTG 144
           ||||||||||||||||||||   ||| |||||||||||| ||||| ||| |||||||||
Db     447 GTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTG 388
```

```
Qy     145 AAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTG 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 AAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTG 328

Qy     205 AAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATG 264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 AAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATG 268

Qy     265 AGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACC 324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 AGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACC 208

Qy     325 CACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATC 384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     207 CACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATC 148

Qy     385 CAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACAC 444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     147 CAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACAC 88

Qy     445 GTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAA 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      87 GTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAA 28

Qy     505 AGCTTTACAGCCTTCTGCAAAGAAAAA 531
           |||||||||||||||||||||||||||
Db      27 AGCTTTACAGCCTTCTGCAAAGAAAAA 1
```

```
RESULT   6
AA738037/c
LOCUS       AA738037                464 bp    mRNA    linear   EST 22-JAN-1998
DEFINITION  nx15e11.s1 NCI_CGAP_GC3 Homo sapiens cDNA clone IMAGE:1256204 3'
            similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;, mRNA sequence.
ACCESSION   AA738037
VERSION     AA738037.1  GI:2768794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 511   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 435.
FEATURES             Location/Qualifiers
     source          1. .464
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1256204"
                     /clone_lib="NCI_CGAP_GC3"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
```

vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 86 a 132 c 93 g 153 t
ORIGIN

Query Match 78.9%; Score 424.4; DB 9; Length 464;
Best Local Similarity 96.9%; Pred. No. 1.9e-88;
Matches 444; Conservative 0; Mismatches 11; Indels 3; Gaps 1;

```
Qy      84 ATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCT 140
           |||||||||||||||||||||||   ||| ||||||| |||||||||| ||| ||||||
Db     464 ATGAGTTGGCGAGGAAGATCGACCTGTTATTGGCCTAGTCCAAGACGCTATGTACAGCCT 405

Qy     141 CCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 200
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     404 CCTGAAATGATTGGGCCTATGCAGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 345

Qy     201 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 285

Qy     261 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGT 320
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     284 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAGGAACAGGGT 225

Qy     321 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 165

Qy     381 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 105

Qy     441 ACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 500
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104 GCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 45

Qy     501 ATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||   |||||||||||||||| ||||||||||||
Db      44 ATAAGAGTTTACAGCCTTCTGCAAAAAAAAAAAAAAA 7
```

RESULT 7
AW016546/c
LOCUS AW016546 418 bp mRNA linear EST 10-SEP-1999
DEFINITION UI-H-BI0p-abg-g-06-0-UI.s1 NCI_CGAP_Sub2 Homo sapiens cDNA clone IMAGE:2711986 3', mRNA sequence.
ACCESSION AW016546
VERSION AW016546.1 GI:5865303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 418)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .418
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2711986"
/clone_lib="NCI_CGAP_Sub2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NCI_CGAP_Sub2 library is a subtracted library derived from BI. BI constitutes a mixture of 21 normalized or subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_CLL1, NCI_CGAP_Lei2, NCI_CGAP_Brn23, NCI_CGAP_Lu5, NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6, NCI_CGAP_Brn25. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition is detailed below: NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs 1322376-1323911, 1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs 1323912-1325831, 1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE CloneIDs 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs 1257096-1258631, 1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759, 1101192-1101959, 1217928-1220615) NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255, 1144584-1145351) The resulting subtracted library contained 4 million recombinants. Subtraction was performed as previously described [Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_LIB=NCI_CGAP_GC4
TAG_TISSUE=germ cell
TAG_SEQ=AAATC"
BASE COUNT 74 a 119 c 87 g 138 t
ORIGIN

Query Match 76.9%; Score 413.8; DB 9; Length 418;
Best Local Similarity 99.5%; Pred. No. 5.5e-86;
Matches 415; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     122 AAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 181
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     418 AAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 359

Qy     182 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 241
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     358 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 299

Qy     242 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 301
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     298 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 239

Qy     302 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 361
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     238 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 179

Qy     362 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 421
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     178 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 119

Qy     422 ATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 481
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     118 ATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 59

Qy     482 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     58 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 2


RESULT   8
BG120336
LOCUS       BG120336                 480 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602353732F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451958 5',
            mRNA sequence.
ACCESSION   BG120336
VERSION     BG120336.1  GI:12613845
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10239  row: p  column: 07
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4451958"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      170 a     93 c    133 g     84 t
ORIGIN


  Query Match           76.8%;  Score 413.4;  DB 10;  Length 480;
  Best Local Similarity  97.4%;  Pred. No. 6.8e-86;
  Matches  420;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    108 TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCC 167
          ||| |||||||||||||| ||||| ||| |||||||||| ||||||||||||||||||||
Db     15 TATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCC 74

Qy    168 GAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAA 227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 GAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAA 134

Qy    228 CGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGG 287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 CGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGG 194

Qy    288 CCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAA 347
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    195 CCGAAGCCTGAAGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAA 254

Qy    348 GATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAA 407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAA 314

Qy    408 GAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCC 467
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 GAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCC 374

Qy    468 TATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGA 527
          ||||||||||||||||||||||||||||||||||||||||||||||||||||  || | |
Db    375 TATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTTGCAAAA 434

Qy    528 AAAAAAAAAAA 538
          ||||||||||| |
Db    435 AAAAAAAAAAA 445


RESULT   9
AA868226/c
LOCUS       AA868226                 426 bp    mRNA    linear   EST 16-MAR-1998
DEFINITION  ak48h07.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1409245
            3' similar to SW:GGE2_HUMAN Q13066 GAGE-2 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AA868226
VERSION     AA868226.1  GI:2963671
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1409245"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']."
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       79 a    119 c     95 g    131 t      2 others
ORIGIN


  Query Match           76.7%;  Score 412.8;  DB 9;  Length 426;
  Best Local Similarity  97.9%;  Pred. No. 9.4e-86;
  Matches  417;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    102 TCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATG 161
          ||||||||||||||||||||||||||||||||||||||||||||||||     |||||||
Db    426 TCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGNANTTGCCTATG 367
```

```
Qy    162 CGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCA 221
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 CGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCA 307

Qy    222 ACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGT 281
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 ACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGT 247

Qy    282 CAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAG 341
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAG 187

Qy    342 TGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACG 401
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 TGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACG 127

Qy    402 CCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGC 461
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    126 CCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGC 67

Qy    462 TGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTG 521
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     66 TGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAACTTTACAGCCTTCTT 7

Qy    522 CAAAGA 527
          |||| |
Db      6 CAAAAA 1
```

```
RESULT  10
AA760996/c
LOCUS       AA760996                450 bp    mRNA    linear   EST 26-JAN-1998
DEFINITION  nx32h08.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1257855 3'
            similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;, mRNA sequence.
ACCESSION   AA760996
VERSION     AA760996.1  GI:2809926
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 450)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 331.
FEATURES             Location/Qualifiers
     source          1. .450
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1257855"
                     /clone_lib="NCI_CGAP_GC4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       85 a    130 c     94 g    140 t      1 others
ORIGIN
```

```
  Query Match           76.4%;  Score 410.8;  DB 9;  Length 450;
  Best Local Similarity  97.3%;  Pred. No. 2.7e-85;
  Matches  439;  Conservative    0;  Mismatches    8;  Indels    4;  Gaps    2;

Qy     84 ATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCT 140
          ||||||||||||||||||||||||   ||| |||||||||||||| ||||| ||| |||||
Db    450 ATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCT 391

Qy    141 CCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 200
          |||||| ||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db    390 CCTGAAGTGANTGTGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 331

Qy    201 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    330 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 271

Qy    261 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGT 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    270 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGT 211

Qy    321 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 151

Qy    381 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 91

Qy    441 ACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90 ACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 31

Qy    501 ATAAAGCTTTACAGCCTTCTGCAAAGAAAAA 531
          |||||| |||||||||||||||| |||||
Db     30 ATAAAG-TTTACAGCCTTCTGCAAAAAAAAA 1
```

```
RESULT  11
AI187350/c
LOCUS       AI187350                455 bp    mRNA    linear   EST 10-NOV-1998
DEFINITION  qf29a05.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1751408
            3' similar to SW:GGE4_HUMAN Q13068 GAGE-4 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AI187350
VERSION     AI187350.1  GI:3737988
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 504   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 391.
FEATURES             Location/Qualifiers
```

```
    source          1. .455
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:1751408"
                    /clone_lib="Soares_testis_NHT"
                    /sex="male"
                    /lab_host="DH10B"
                    /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                    was prepared from mRNA obtained from Clontech Laboratories
                    , Inc., and primed with  a Not I - oligo(dT) primer [5'
                    TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                    Double-stranded cDNA was ligated to Eco RI adaptors
                    (Pharmacia), digested with Not I and cloned into the Not I
                    and Eco RI sites of the modified pT7T3 vector. Library
                    went through one round of normalization to Cot5, and was
                    constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       86 a    130 c     95 g    144 t
ORIGIN


  Query Match             74.6%;  Score 401.2;  DB 9;  Length 455;
  Best Local Similarity   95.4%;  Pred. No. 4.6e-83;
  Matches  435;  Conservative    0;  Mismatches   18;  Indels    3;  Gaps    2;

Qy      84 ATGAGTTGGCGAGGAAGATCGACCTAT--CGGCCTAGACCAAGACGCTACGTAGAGCCTC 141
           ||||||||||||||||||    | |||    | |     |||| ||||| ||  ||||||
Db     455 ATGAGTTGGCGAGGAAGATGACCTTATTATTGGCCTAGACAAGGCGCTATGT-CAGCCTC 397

Qy     142 CTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACAC 201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     396 CTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACAC 337

Qy     202 CTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGG 261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     336 CTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGG 277

Qy     262 ATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTC 321
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     276 ATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTC 217

Qy     322 ACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAA 381
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     216 ACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAA 157

Qy     382 ATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGA 441
           |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     156 ATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCCCAGTGTTAAAAGAAGA 97

Qy     442 CACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAA 501
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      96 CACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAA 37

Qy     502 TAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
           ||||||||| |||||||||||||| ||||||||||
Db      36 TAAAGCTTTCCAGCCTTCTGCAAAAAAAAAAAAAAA 1


RESULT  12
AW102587/c
LOCUS       AW102587                412 bp    mRNA    linear   EST 19-OCT-1999
DEFINITION  xd67g07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2602716 3' similar to SW:GGE2_HUMAN Q13066 GAGE-2 PROTEIN. [1
            ] ;, mRNA sequence.
ACCESSION   AW102587
VERSION     AW102587.1  GI:6073200
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 412)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco
            High quality sequence stop: 285.
FEATURES             Location/Qualifiers
     source          1. .412
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2602716"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       78 a    114 c     84 g    133 t      3 others
ORIGIN


  Query Match             67.9%;  Score 365.4;  DB 9;  Length 412;
  Best Local Similarity   97.8%;  Pred. No. 9.6e-75;
  Matches  391;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy     138 CCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCA-GC 196
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     400 CCTCTTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCANGC 341

Qy     197 AACACCTGAAGAAGGGGAACCAG-CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGG 255
           ||||| |||||||||||||||| ||||||| |||||||||| |||||||||| |||
Db     340 AACACTTGAAGAAGGGGAACCAGNCAACTCANAGTCAGGATCCTTCAGCTGCTCAGAAGG 281

Qy     256 GAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAAC 315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     280 GAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAAC 221

Qy     316 AGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACC 375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     220 AGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACC 161

Qy     376 CGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAA 435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     160 CGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAA 101

Qy     436 AGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTC 495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100 AGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTC 41

Qy     496 TCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAA 535
           |||||||||||||||||||||||||||| |||||||||
Db      40 TCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAA 1


RESULT  13
BC005363
LOCUS       BC005363                445 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, clone IMAGE:3686872, mRNA.
ACCESSION   BC005363
VERSION     BC005363.1  GI:14710126
KEYWORDS    HTC.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: e Column: 15
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503882
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .445
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3686872"
                     /tissue_type="Bone marrow, chronic myelogenous leukemia"
                     /clone_lib="NIH_MGC_54"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
BASE COUNT      127 a    102 c    132 g     84 t
ORIGIN
```

```
  Query Match             67.0%;  Score 360.4;  DB 11;  Length 445;
  Best Local Similarity   95.3%;  Pred. No. 1.4e-73;
  Matches  383;  Conservative    0;  Mismatches   16;  Indels    3;  Gaps    1;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 94

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||
Db      95 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTGTTATTGGCCTA 154

Qy     118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
           | |||||||||||| ||| ||||||||||||||||||||||||||| ||||||||||||||
Db     155 GTCCAAGACGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCAGCCCGAGCAGTTCA 214

Qy     178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 274

Qy     238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 334

Qy     298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
           ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 AAGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 394

Qy     358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 399
           |||||||||||||||||||||||||   | |  |    ||||
Db     395 ATGGGCAGGAGATGGACCCGCCAAAAAAAAAAAAAAAAAAAA 436
```

```
RESULT  14
AA448542
LOCUS       AA448542                 419 bp    mRNA    linear   EST 04-JUN-1997
DEFINITION  zw81e11.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782636
            5' similar to TR:G914905 G914905 GAGE-4 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AA448542
VERSION     AA448542.1  GI:2162212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 419)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 353.
FEATURES             Location/Qualifiers
     source          1. .419
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:782636"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. "
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      108 a     98 c    134 g     79 t
ORIGIN
```

```
  Query Match             66.8%;  Score 359.4;  DB 9;  Length 419;
  Best Local Similarity   96.3%;  Pred. No. 2.4e-73;
  Matches  390;  Conservative    0;  Mismatches   11;  Indels    4;  Gaps    2;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 75

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||
Db      76 ACTGAGATTCATCTGTGTGAAATATGAGTAGGCGAGGAAGATCGACCTATTATTGGCCTA 135

Qy     118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
           ||||||| ||||| ||| |||||||||| |||||||||||||||||| ||||||||||
Db     136 GACCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCAGAGCAGTTCA 195

Qy     178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     196 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 255
```

```
Qy    238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
          ||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db    256 CTGCAGCTGCTCATCGCGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 315

Qy    298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
          |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    316 AAGCTCATAGCCAGGAACAGGGTCA-CCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 374

Qy    358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGC 402
          |||||||||||||||||||||||||||||||||||||||||||||
Db    375 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGC 419


RESULT 15
AA913206/c
LOCUS       AA913206                384 bp    mRNA    linear   EST 17-JUN-1998
DEFINITION  om53f03.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1550909 3'
            similar to SW:GGE3_HUMAN Q13067 GAGE-3 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AA913206
VERSION     AA913206.1  GI:3052598
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 487   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 81.
FEATURES             Location/Qualifiers
     source          1. .384
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1550909"
                     /clone_lib="NCI_CGAP_GC4"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is normalized. Library was constructed by
                     Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       63 a    119 c     83 g    119 t
ORIGIN


  Query Match           61.3%;  Score 329.8;  DB 9;  Length 384;
  Best Local Similarity   94.0%;  Pred. No. 1.8e-66;
  Matches  343;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    100 GATCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTA 159
          || | |  ||| |||||||||||||| ||||| ||| ||||||||||||       |||||
Db    365 GACCTATTTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAAAATGAATTGCCTA 306

Qy    160 TGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 219
          |||| || || |||||||| ||||||||||||||||||||||||||||||||||||||||
Db    305 TGCGCCCGGAACAGTTCAGGGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 246

Qy    220 CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 279
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    245 CAACTCAACGTCAGAATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 186

Qy    280 GTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 339
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 GTCAAGTGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 126

Qy    340 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 66

Qy    400 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAG 459
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db     65 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAG 6

Qy    460 GCTGC 464
          |||||
Db      5 GCTGC 1


Search completed: July  1, 2002, 07:07:18
Job time: 1784 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:39:24 ; Search time 177.44 Seconds
(without alignments)
5205.703 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 2825752

Minimum DB seq length: 0
Maximum DB seq length: 538

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	538	100.0	538	20	AAX90519	GAGE-2 tumour reje
	2	504.8	93.8	532	20	AAX90522	GAGE-5 tumour reje
	3	496.8	92.3	532	19	AAV18720	cDNA encoding GAGE
	4	493	91.6	535	19	AAV18717	cDNA encoding GAGE
	5	357.8	66.5	430	21	AAC02129	Human secreted pro
c	6	301.4	56.0	530	22	AAS60104	Human cancer agent
c	7	299.2	55.6	365	22	AAS60496	Human cancer agent
	8	194	36.1	532	22	AAI60530	Human polynucleoti
	9	172.4	32.0	503	23	AAS69484	DNA encoding novel
	10	130.4	24.2	475	22	AAD14981	Human NOV2 DNA. H
	11	108	20.1	399	22	AAF68151	Human lung tumour
	12	101.6	18.9	515	22	AAS63900	Human prostate cDN
	13	101.6	18.9	515	22	AAH93807	Human prostate-spe
	14	101.6	18.9	515	22	AAH85121	Human prostate-spe
	15	101.6	18.9	515	22	AAH02872	Prostate tumour an
	16	100.8	18.7	214	22	AAS24637	Human ovarian PCR-
	17	100.8	18.7	320	22	AAH83260	Human ovarian tumo
c	18	93.4	17.4	472	21	AAZ95012	CSG Pro118 EST use
	19	93.4	17.4	476	21	AAZ93524	PAGE-4 coding sequ
	20	93.4	17.4	478	21	AAZ95013	Cancer specific ge
	21	92.8	17.2	457	22	AAF68851	Human lung tumour
	22	92.4	17.2	420	21	AAA06691	Human immunogenic
c	23	86.4	16.1	348	22	AAD14997	pCR2.1-83350421-S7
	24	84.8	15.8	479	22	AAF68152	Human lung tumour
	25	82.8	15.4	458	22	AAD14980	Human NOV1 DNA. H
	26	79.2	14.7	372	20	AAX41280	Human secreted pro
	27	77.4	14.4	461	22	AAF68852	Human lung tumour
	28	75.4	14.0	328	20	AAV87256	EST clone BN365.
	29	68.4	12.7	300	20	AAX55997	Human cDNA clone F
c	30	53	9.9	157	21	AAC15593	Human secreted pro
c	31	53	9.9	201	21	AAA42613	Human secreted exp
	32	51	9.5	51	22	AAH79806	Human DNA containi
	33	49.4	9.2	51	22	AAL33515	Human SNP oligonuc
	34	49.4	9.2	51	22	AAH79998	Human DNA containi
	35	46.2	8.6	51	22	AAL32796	Human SNP oligonuc
	36	44.6	8.3	51	22	AAH79997	Human DNA containi
	37	44.2	8.2	51	22	AAL27748	Human SNP oligonuc
	38	40	7.4	372	23	AAS75452	DNA encoding novel
	39	38.2	7.1	51	22	AAL27747	Human SNP oligonuc
	40	38	7.1	51	22	AAL32276	Human SNP oligonuc
	41	37.8	7.0	498	21	AAC45680	Arabidopsis thalia
	42	37.8	7.0	500	21	AAC39188	Arabidopsis thalia
	43	37.4	7.0	50	22	AAL32275	Human SNP oligonuc
	44	37.2	6.9	237	21	AAC66318	EIAV tat gene SEQ
	45	36.6	6.8	501	23	AAS66773	DNA encoding novel

ALIGNMENTS

RESULT 1
AAX90519
ID AAX90519 standard; cDNA; 538 BP.
XX
AC AAX90519;
XX
DT 30-SEP-1999 (first entry)
XX
DE GAGE-2 tumour rejection antigen clone nucleotide sequence.
XX
KW Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW detection; therapy; pathological condition; cancer; CTL;
KW cytolytic T lymphocyte; GAGE; ss.
XX
OS Homo sapiens.
XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX
DR WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of

```
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
SQ   Sequence 538 BP; 160 A; 116 C; 155 G; 107 T; 0 other;


  Query Match             100.0%;  Score 538;  DB 20;  Length 538;
  Best Local Similarity   100.0%;  Pred. No. 3e-146;
  Matches  538;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 acgccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttttcctct 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 actgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctatcggcctagac 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 caagacgctacgtagagcctcctgaaatgattgggcctatgcggcccgagcagttcagtg 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 atgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctg 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 cagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaag 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctcatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatg 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagc 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aatcacagtgttaaaagaagacacgttgaaatgatgcaggctgctcctatgttggaaatt 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tgttcattaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 538


RESULT   2
AAX90522
ID   AAX90522 standard; cDNA; 532 BP.
XX
AC   AAX90522;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE-5 tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937665-A1.
XX
PD   29-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00775.
XX
PR   23-JAN-1998;   98US-0012818.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B,  Van Der Bruggen P;
XX
DR   WPI; 1999-469111/39.
XX
PT   New isolated peptides which bind to HLA-A29 molecules, which are
PT   tumour rejection antigens used for detection and therapy of
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
SQ   Sequence 532 BP; 156 A; 111 C; 154 G; 111 T; 0 other;


  Query Match             93.8%;  Score 504.8;  DB 20;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 1.3e-136;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agctgtgaggcagtgctgtgtggttcctgccgtccggactcttttttcctctactgagatt 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           ||||||||||||||||||||||||||||||||||||   ||| ||||||||||||| |
Db      61 catctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctagaccaaggc 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     121 gctatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagtgatgaag 180

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctgcagctg 240

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     241 ctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaagctgata 300

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatgggcagg 360

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 agatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagcaatcac 420

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     421 agtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatttgttca 480

Qy     487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ttaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 532
```

```
RESULT   3
AAV18720
ID   AAV18720 standard; cDNA; 532 BP.
XX
AC   AAV18720;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-5 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             75..429
FT                   /*tag= a
FT                   /transl_except= (pos:189..191, aa:Ala)
FT                   /transl_except= (pos:192..194, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47602.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-5 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 532 BP; 154 A; 110 C; 156 G; 112 T; 0 other;


  Query Match             92.3%;  Score 496.8;  DB 19;  Length 532;
  Best Local Similarity   97.2%;  Pred. No. 2.7e-134;
  Matches  517;  Conservative    0;  Mismatches   12;  Indels    3;  Gaps    1;

Qy     10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 agctgtgaggcagtgctgtgtggttcctgccgtccggactctttttcctctactgagatt 60

Qy     70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           ||||||||||||||||||||||||||||||||||||   ||| ||||||||||||| |

Db     61 catctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctagaccaaggc 120

Qy    127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           | || ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    121 ggtatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagtgatgaag 180

Qy    187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           |||||||||    | ||||||||||||||||||||||||||||||||||||||||||||
Db    181 tggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcctgcagctg 240

Qy    247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    241 ctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaagctgata 300

Qy    307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatgggcagg 360

Qy    367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 agatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagcaatcac 420

Qy    427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    421 agtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatttgttca 480

Qy    487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ttaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 532


RESULT   4
AAV18717
ID   AAV18717 standard; cDNA; 535 BP.
XX
AC   AAV18717;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-2 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             81..431
FT                   /*tag= a
FT                   /transl_except= (pos:192..194, aa:Ala)
FT                   /transl_except= (pos:195..197, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47599.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
```

```
XX
CC   The present sequence encodes a GAGE-2 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 535 BP; 158 A; 112 C; 156 G; 109 T; 0 other;


  Query Match             91.6%;  Score 493;  DB 19;  Length 535;
  Best Local Similarity   97.2%;  Pred. No. 3.5e-133;
  Matches  524;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    2;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTC- 59
           |||||||||||||||||||||||||||||    | ||  ||||||||||||||||||||
Db       1 acgccagggagctgtgaggcagtgctgtg----ttcttgcgtccggactcttttcctct 56

Qy      60 TACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGA 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57 tactgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctatcggcctaga 116

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117 ccaagacgctacgtagagcctcctgaaatgattgggcctatgcggcccgagcagttcagt 176

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||   | ||||||||||||||||||||||||||||||||||||||
Db     177 gatgaagtggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcct 236

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 gcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaa 296

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 gctgagagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgat 356

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 gggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaag 416

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     417 caatcacagtgttaaaagaagacatgttgaaatgatgcaggctgctcctatgttggaaat 476

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     477 ttgttcattaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 535


RESULT   5
AAC02129
ID   AAC02129 standard; cDNA; 430 BP.
XX
AC   AAC02129;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 2127.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG02123.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 2127; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 430 BP; 101 A; 102 C; 134 G; 85 T; 8 other;


  Query Match             66.5%;  Score 357.8;  DB 21;  Length 430;
  Best Local Similarity   97.6%;  Pred. No. 5.9e-94;
  Matches  368;  Conservative    5;  Mismatches    2;  Indels    2;  Gaps    1;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      54 acgccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctct 113

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTAT--CGGCCTAG 118
           |||||||||||||||||||||||||||||||||||||||||||||||||  :|||||||
Db     114 actgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctatnayggcctag 173

Qy     119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
           ||||||:||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db     174 accaagrcgctatgtacagcctcctgaaatgattgggcctatgcggcccgagcagttcag 233

Qy     179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
           |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db     234 tgatgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaaygtcaggatcc 293

Qy     239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
           |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db     294 tgcagctgctcagragggagaggatgagggagcatctgcaggtcaagggccgaagcctga 353

Qy     299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     354 agctcatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctga 413

Qy     359 TGGGCAGGAGATGGACC 375
           |||||||||||||||:|
```

```
Db    414 tgggcaggagatggamc 430


RESULT   6
AAS60104/c
ID   AAS60104 standard; cDNA; 530 BP.
XX
AC   AAS60104;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human cancer agent-sensitive marker #105.
XX
KW   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW   Hodgkin's disease; glioma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179556-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US12132.
XX
PR   14-APR-2000; 2000US-197538P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Brown JL,  Bolt A,  Van Huffel C;
XX
DR   WPI; 2001-602933/68.
XX
PT   Novel nucleic acid, used as a marker to determine the effectiveness of
PT   using TAXOL to treat cancer cell growth in individuals -
XX
PS   Claim 1; Page 120; 527pp; English.
XX
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC   they are shown to express one of the 242 sensitivity markers or the
CC   cells are shown not to express one of the 804 resistance markers.
CC   The methods can be used to determine the effectiveness of TAXOL
CC   in the treatment of cancer cell growth in an individual. The markers
CC   can be used as targets in developing anti-cancer agents such as
CC   chemotherapeutic compounds. The markers can also be used as targets in
CC   developing treatments for cancer, particularly those cancers which
CC   display resistance to agents and exhibit expression of the markers. The
CC   anticancer agents developed by the novel method can be used to treat
CC   cancer. Probes based on the markers can be used to detect transcripts or
CC   genomic sequences corresponding to the markers, in the identification of
CC   cells or tissues which mis-express the protein. Cancers which may
CC   be targeted include carcinoma (e.g. squamous cell carcinoma),
CC   sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
CC   lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
CC   tumours (e.g. glioma). The present sequence is one of the 1046
CC   novel cancer cell markers.
XX
SQ   Sequence 530 BP; 89 A; 100 C; 106 G; 183 T; 52 other;


  Query Match             56.0%;  Score 301.4;  DB 22;  Length 530;
  Best Local Similarity   87.9%;  Pred. No. 1.6e-77;
  Matches  320;  Conservative    0;  Mismatches   42;  Indels    2;  Gaps    1;

Qy    177 AGTGATGAAGTGGAACCAGC--AACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGG 234
          ||| || ||||||| ||| |  |   |||||| ||||| |||||||||||||||||| |||
Db    449 AGTNATNAAGTGGACCCANCCNACNCCCTGAANAAGGGNAACCAGCAACTCAACGTNAGG 390

Qy    235 ATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGC 294
          |||||| |||||| ||| ||||| ||||| |||| |||| ||||||||||||||||||||

Db    389 ATCCTGNAGCTGCNCAGNAGGGANAGGATNAGGGNGCATNTGCAGGTCAAGGGCCGAAGC 330

Qy    295 CTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTC 354
          |||||| | |||||||||||||||| ||| || |||||| || ||||| || ||||||
Db    329 CTGAAGNTGATAGCCAGGAACAGGGTCNCCCCCANACTGGGNGTNAGTGTNAAAATGGTC 270

Qy    355 CTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTG 414
          || |||||||| | |||||||| ||||||||| || |||||||||| |||||| ||||||
Db    269 CTNATGGGCAGNANATGGACCCNCCAAATCCAAAGNAGGTGAAAACNCCTGAANAAGGTG 210

Qy    415 AAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTG 474
          |||||||||| |||||||||| ||| | ||||||||||||||||||||||||||||||
Db    209 AAAAGCAATCCCAGTGTTAAAANAAGGCNCGTTGAAATGATGCAGGCTGCTCCTATGTTG 150

Qy    475 GAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAA 534
          |||||||||||||||||||||||||||||||||||||||||||||| |  ||| ||||||||
Db    149 GAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTNTNAAAAAAAAAAAAA 90

Qy    535 AAAA 538
          ||||
Db     89 AAAA 86


RESULT   7
AAS60496/c
ID   AAS60496 standard; cDNA; 365 BP.
XX
AC   AAS60496;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human cancer agent-sensitive marker #227.
XX
KW   Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW   squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW   lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW   Hodgkin's disease; glioma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200179556-A2.
XX
PD   25-OCT-2001.
XX
PF   13-APR-2001; 2001WO-US12132.
XX
PR   14-APR-2000; 2000US-197538P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Brown JL,  Bolt A,  Van Huffel C;
XX
DR   WPI; 2001-602933/68.
XX
PT   Novel nucleic acid, used as a marker to determine the effectiveness of
PT   using TAXOL to treat cancer cell growth in individuals -
XX
PS   Claim 1; Page 231; 527pp; English.
XX
CC   The invention relates to 1046 novel nucleic acids which are used as
CC   markers for determining the sensitivity of a cancer cell to the
CC   anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC   they are shown to express one of the 242 sensitivity markers or the
CC   cells are shown not to express one of the 804 resistance markers.
CC   The methods can be used to determine the effectiveness of TAXOL
CC   in the treatment of cancer cell growth in an individual. The markers
CC   can be used as targets in developing anti-cancer agents such as
CC   chemotherapeutic compounds. The markers can also be used as targets in
CC   developing treatments for cancer, particularly those cancers which
CC   display resistance to agents and exhibit expression of the markers. The
CC   anticancer agents developed by the novel method can be used to treat
CC   cancer. Probes based on the markers can be used to detect transcripts or
CC   genomic sequences corresponding to the markers, in the identification of
```

```
CC   cells or tissues which mis-express the protein. Cancers which may
CC   be targeted include carcinoma (e.g. squamous cell carcinoma),
CC   sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
CC   lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
CC   tumours (e.g. glioma). The present sequence is one of the 1046
CC   novel cancer cell markers.
XX
SQ   Sequence 365 BP; 63 A; 86 C; 66 G; 150 T; 0 other;


  Query Match             55.6%;  Score 299.2;  DB 22;  Length 365;
  Best Local Similarity   95.9%;  Pred. No. 5.7e-77;
  Matches  307;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy     219 GCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCA 278
           ||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||||
Db     365 GCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAAAGGATAAGGGAGCATCTGCA 306

Qy     279 GGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGT 338
           |||||||||||||||||||||| |||||||||||||||||||| ||||||||||||
Db     305 GGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAGGAACAGGGTCACCCCCAGACTGGGTGT 246

Qy     339 GAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAA 398
           ||||||||| ||||||||||||||||||| |||||||||||||||| ||||||||||||
Db     245 GAGTGTGAAAATGGTCCTGATGGGCAGGAAATGGACCCGCCAAATCCAAAGGAGGTGAAA 186

Qy     399 ACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCA 458
           |||||||||||||||||||||||||| |||||||||| ||| ||||||||||||||||
Db     185 ACGCCTGAAGAAGGTGAAAAGCAATCTCAGTGTTAAAAAAAGGCACGTTGAAATGATGCA 126

Qy     459 GGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTT 518
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125 GGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTT 66

Qy     519 CTGCAAAGAAAAAAAAAAAA 538
           ||  ||| ||||||||||||
Db      65 CTAAAAAAAAAAAAAAAAAA 46


RESULT   8
AAI60530
ID   AAI60530 standard; cDNA; 532 BP.
XX
AC   AAI60530;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4519.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia; ss.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   P-PSDB; AAM41374.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Claim 1; SEQ ID NO 4519; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence 532 BP; 154 A; 108 C; 150 G; 120 T; 0 other;


  Query Match             36.1%;  Score 194;  DB 22;  Length 532;
  Best Local Similarity   68.4%;  Pred. No. 2.4e-46;
  Matches  365;  Conservative    0;  Mismatches  145;  Indels   24;  Gaps    6;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTG-CTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTC 59
           | |||  |||| |||||||  |||   ||   |||||||| ||| ||||| ||| |   |
Db      18 atgccctggagttgtgagggtgtgagggtcgcgttcctgctgtctggacttttttctgtcc 77

Qy      60 TACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGA 119
           |||||||  || ||||||||||||||| |||||||||||||| || ||| |||||||
Db      78 cactgagacgcagctgtgtgaaatatgatttggcgaggaagatcaacatataggcctagg 137

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           || ||  |   |||  |||||||| |||||||||||||||       | ||  | ||
Db     138 ccgaggagaagtgtaccacctcctgagctgattgggcctatgc------tggagcccggt 191

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           |||||  | |            ||| |  ||| |||||||| ||||| || ||| |||||||
Db     192 gatgaggag-----------cctcagcaagaggaaccaccaactgaaagtcgggatcct 239

Qy     240 GCAGCTGCTCAGGAG---GGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCT 296
           ||| ||| |||||||   | ||| ||| |||| ||| |||    ||||| |||  | |
Db     240 gcacctggtcaggagagagaagaagatcagggtgcagctgagactcaagtgcctgacctg 299

Qy     297 GAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCT 356
           |||||| ||  |||||| | |  ||| ||  |||||||| ||| || | ||||||||||
Db     300 gaagctgatctccaggagctgtctcagtcaaagactggggatgaatgcggagatggtcct 359

Qy     357 GATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAA-GAAGGTGA 415
           ||||  |||| || |   | | |||||  |||| |  | |||| ||| ||| | ||||||
Db     360 gatgtccaggggaagattctgacaaagtcagagcaatttaaaatgccagaagggaggtga 419

Qy     416 AAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGG 475
           | |||| |||||  ||||| |||||||| | |||||  |  ||      |  | | |
Db     420 caggcaaccacaggtttaaatgaagacaagctgaaacaacacaaaactgt-ttttatcta 478

Qy     476 AAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAA 529
           | || | |   |||||| | ||  ||||||  ||| |||| |||| | || |||
Db     479 agatatttgacttaaaaatatcgaaataaacttttgcagctttctccgaaaaaa 532
```

```
RESULT   9
AAS69484
ID   AAS69484 standard; cDNA; 503 BP.
XX
AC   AAS69484;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #5288.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG05297.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 5288; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 503 BP; 158 A; 98 C; 121 G; 105 T; 21 other;


  Query Match             32.0%;  Score 172.4;  DB 23;  Length 503;
  Best Local Similarity   68.2%;  Pred. No. 4.4e-40;
  Matches  307;  Conservative    0;  Mismatches  121;  Indels   22;  Gaps    4;

Qy      76 TGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTACGTAG 135
           |||||||||||| |||||||||||||| || ||| ||||||| || ||  |    |||
Db      52 tgtgaaatatgatttggcgaggaagatcaacatataggcctaggccgaggagaagtgtac 111

Qy     136 AGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAG 195
           ||||||||  ||||||||||||||||        | ||  | |||||||| | |
Db     112 cacctcctgagctgattgggcctatgc------tggagcccggtgatgaggag------- 158

Qy     196 CAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGG 255
               ||| |  ||| ||||||| ||||| || ||| |||||||||| ||| |||||||
Db     159 -----cctcagcaagaggaaccaccaactgaaagtcgggatcctgcacctggtcaggaga 213

Qy     256 G---AGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGG 312
           |   ||| ||| |||||  || ||     ||||| |||  | |  |||||| ||  |||||
Db     214 gaaaagaagatcagggttcagctaagactcaagtgcctgacctggaagctgatctccagg 273

Qy     313 AACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGG 372
           | | |  |||  || |||||||| |||| |||| | ||||||||||||  |||| || |
Db     274 agctgtctcagtcaaagactgggggtgaatgtggaaatggtcctgatgaccaggggaaga 333

Qy     373 ACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTT 432
             | ||||||  ||||  |  | |||| ||| |||| |||||| | |||| |||||  ||
Db     334 ttctgccaaaatcagaacaatttaaaatgccagaaggaggtgacaggcaaccacaggttt 393

Qy     433 AAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAA 492
           ||| ||||||| | |||||  |  ||       |  | | ||  | || | |   ||||||
Db     394 aaatgaagacaagctgaaacaacccaaaactgt-ttttatttaagatatttgacttaaaa 452

Qy     493 TTCTCCCAATAAAGCTTTACAGCCTTCTGC 522
           | ||  ||||||  ||| |||| |||| |
Db     453 atatcgaaataaacttttgcagctttctcc 482


RESULT  10
AAD14981
ID   AAD14981 standard; DNA; 475 BP.
XX
AC   AAD14981;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Human NOV2 DNA.
XX
KW   Human; NOVX; G-antigen; GAGE-like protein; interferon;
KW   G-protein coupled receptor; GPCR; hepatocyte nuclear factor;
KW   mast cell protease; gene therapy; proliferative disorder; cancer;
KW   immune disorder; hepatic disorder; cirrhosis; viral infection;
KW   hepatitis; neuroolfactory system-related disorder; neurological disorder;
KW   Parkinson's disease; infertility; autoimmune disease; arthritis;
KW   multiple sclerosis; allergy; wound healing; cytostatic; nootropic;
KW   immunosuppressive; neuroprotective; vulnerary; hepatotropic; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..24
FT                   /*tag=  a
FT   CDS             25..360
FT                   /*tag=  b
FT                   /product= "Human NOV2 protein"
FT   3'UTR           361..475
FT                   /*tag=  c
XX
PN   WO200161009-A2.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US04828.
XX
PR   15-FEB-2000; 2000US-0182723.
PR   15-FEB-2000; 2000US-0182724.
PR   15-FEB-2000; 2000US-0182733.
PR   22-FEB-2000; 2000US-0183896.
PR   23-FEB-2000; 2000US-0184275.
PR   23-FEB-2000; 2000US-0184482.
PR   23-FEB-2000; 2000US-0184497.
PR   24-FEB-2000; 2000US-0184744.
```

```
PR   13-APR-2000; 2000US-0197083.
PR   10-AUG-2000; 2000US-0224157.
PR   18-SEP-2000; 2000US-0233405.
PR   27-SEP-2000; 2000US-0236060.
PR   02-JAN-2001; 2001US-0259414.
PR   18-JAN-2001; 2001US-0262454.
PR   14-FEB-2001; 2001US-0783429.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Malyankar UM,  Tchernev VT,  Padigaru M,  Taupier RJ,  Spytek KA;
PI   Majumder K,  Guo X,  Spaderna SK,  Boldog FL;
XX
DR   WPI; 2001-514775/56.
DR   P-PSDB; AAE08581.
XX
PT   Isolated novel polypeptides useful for diagnosis of and treating
PT   cancer, infertility, autoimmune diseases, arthritis, multiple
PT   sclerosis, allergies, wound healing and hepatic disorders -
XX
PS   Claim 9; Page 8; 140pp; English.
XX
CC   The present sequence is a human NOV2 DNA. The NOVX protein has homology
CC   with one of G-antigen (GAGE)-like protein, interferon, G-protein coupled
CC   receptor (GPCR), hepatocyte nuclear factor or mast cell protease. The
CC   NOVX is useful for treating or preventing a pathology associated with
CC   NOVX. It is also useful for determining the presence or amount of NOVX
CC   DNA in a sample, for identifying a potential therapeutic agent and in
CC   gene therapy. It is also useful for determining the presence of or
CC   predisposition to a disease associated with altered levels of NOVX. It is
CC   also useful for the diagnosis and treatment of proliferative disorders,
CC   e.g., cancer, immune disorders, hepatic disorders, e.g., cirrhosis, viral
CC   infections, e.g., hepatitis, neuroolfactory system-related disorders,
CC   neurological disorders, e.g., Parkinson's disease, infertility,
CC   autoimmune diseases, arthritis, multiple sclerosis, allergies and wound
CC   healing.
XX
SQ   Sequence 475 BP; 162 A; 85 C; 122 G; 106 T; 0 other;

  Query Match             24.2%;  Score 130.4;  DB 22;  Length 475;
  Best Local Similarity   64.9%;  Pred. No. 6.8e-28;
  Matches  227;  Conservative    0;  Mismatches  116;  Indels    7;  Gaps    2;

Qy     195 GCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAG 254
           | || || | |||||| ||||||| ||||| |   ||||| |  |||| ||  |  ||||
Db     127 gaaaaacgtcaagaagaggaaccaccaactgataatcagggtattgcacctagtggggag 186

Qy     255 GGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAA 314
             ||  |||| |||||| |||| | |||||||||  |    ||||||  |  ||||||
Db     187 atcgaaaatgaaggagcacctgccgttcaagggcctgacatggaagcttttcaacaggaa 246

Qy     315 CAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGAC 374
           | || ||  |   |||  | |  |||| || ||||||||||||||    ||||  |
Db     247 ctggctctgcttaagatagaggatgagcctggagatggtcctgatgtcagggaggggatt 306

Qy     375 CCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAA 434
             ||| | |   ||       |||  ||  |||| ||||||   |||| || ||  || |
Db     307 atgcccacttttgatctcactaaagtgctggaagcaggtgatgcgcaaccataggtttca 366

Qy     435 AAGAAGACACGT------TGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATT 488
           |  ||||||  |      |||||  | | | | |  || ||   ||||||||||     |
Db     367 agcaagacaaatgaagactgaaaccaagaacgttattcttaatctggaaatttgactgat 426

Qy     489 AAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           || ||||||  ||||||| |||| ||  ||||||||||||||||||||
Db     427 aatattctcttaataaagtttta-agttttctgcaaagaaaaaaaaaaaa 475


RESULT  11
AAF68151
ID   AAF68151 standard; cDNA; 399 BP.
XX
AC   AAF68151;
XX
DT   12-APR-2001  (first entry)
XX
DE   Human lung tumour protein related nucleotide sequence SEQ ID NO:69.
XX
KW   Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
KW   lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
KW   cytostatic; antisense inhibition; ss.
XX
OS   Homo sapiens.
XX
PN   WO200100828-A2.
XX
PD   04-JAN-2001.
XX
PF   30-JUN-2000; 2000WO-US18061.
XX
PR   30-JUN-1999;   99US-0346492.
PR   15-OCT-1999;   99US-0419356.
PR   17-DEC-1999;   99US-0466867.
PR   30-DEC-1999;   99US-0476300.
PR   06-MAR-2000; 2000US-0519642.
PR   22-MAR-2000; 2000US-0533077.
PR   10-APR-2000; 2000US-0546259.
PR   27-APR-2000; 2000US-0560406.
PR   05-JUN-2000; 2000US-0589184.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Wang T,  Bangur CS,  Lodes MJ,  Fanger GR,  Vedvick TS,  Carter D;
PI   Retter MW,  Mannion J;
XX
DR   WPI; 2001-071488/08.
XX
PT   Lung tumor-associated proteins and the nucleic acids that encode them,
PT   useful for preventing, diagnosing and treating lung cancer -
XX
PS   Claim 4; Page 173; 436pp; English.
XX
CC   The present invention describes immunogenic portions of lung tumour-
CC   associated proteins (I) and the nucleic acids (NAs) that encode them.
CC   (I) have cytostatic activity and can be used in gene therapy, antisense
CC   inhibition and in vaccines. The NAs and the lung tumour-associated
CC   proteins they encode may be used in the prevention, treatment and
CC   diagnosis of diseases associated with their inappropriate expression,
CC   especially lung cancers. For example, the NAs may be administered to
CC   treat diseases by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of the protein by expressing inactive proteins
CC   or to supplement the patients own production of (I). Additionally, the
CC   NAs may be used to produce the lung-tumour associated protein, according
CC   to standard recombinant DNA methodology. Conversely, antisense NA
CC   molecules may be administered to down regulate protein expression by
CC   binding with the cells own genes and preventing their expression. The NA
CC   and complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar NA sequences in
CC   samples, and hence which patients may be in need of treatment for lung
CC   cancer. The (I) may be used as antigens in the production of antibodies
CC   and in assays to identify modulators (agonists and antagonists) of the
CC   expression and activity of the protein. AAF68083 to AAF68878 and
CC   AAB76848 to AAB76878 represent human lung tumour protein related
CC   nucleotide and protein sequences which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 399 BP; 150 A; 87 C; 94 G; 67 T; 1 other;

  Query Match             20.1%;  Score 108;  DB 22;  Length 399;
  Best Local Similarity   63.3%;  Pred. No. 2e-21;
  Matches  214;  Conservative    1;  Mismatches  116;  Indels    7;  Gaps    3;

Qy     201 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAG---GGA 257
           ||  || |||  ||||||||| || || ||| ||||||| || |||  |||  ||   | |
```

```
Db      65 ccccaaaaagaagaaccagcagctgaaagtcgggatcctacacctgggcagcagacagaa 124

Qy     258 GAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAG 317
           || ||| |||   || |||     | || ||   |    |||| | ||   || || | |
Db     125 gaagatcaggatacagctgagatcccagtgcgcgacatggaaggtgatctgcaagagctg 184

Qy     318 GGTCACCCACAGACTGGGTGTGAGTGTGAAGAT-GGTCCTGATGGGCAGGAGATGGACCC 376
             |||  || | ||| |||  | | | || |  | ||| | |  |   ||| || |
Db     185 catcagtcaaacaccggggataaatctggatttgggttccggcgtcaaggtgaagataat 244

Qy     377 GCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAA 436
            || ||    ||  |    |||| ||| |||| |||||||| ||||| ||||   |||||
Db     245 acctaaagaggaacactgtaaaatgccagaagcaggtgaagagcaaccacaagtttaaat 304

Qy     437 GAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCT 496
           ||||||| | |||||  | ||| |||| |  ||| || || ||||   | | ||| | |
Db     305 gaagacaagctgaaacaacgcaagctggttttatattagatatttg---acttaaactat 361

Qy     497 CCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAA 534
           | |||||||| ||| |||| |||  |||: ||||||||
Db     362 ctcaataaagttttgcagctttcaccaaraaaaaaaaa 399


RESULT 12
AAS63900
ID   AAS63900 standard; cDNA; 515 BP.
XX
AC   AAS63900;
XX
DT   29-JAN-2002  (first entry)
XX
DE   Human prostate cDNA sequence #434.
XX
KW   Human; prostate cancer; ss; cytostatic; immunostimulant; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2001-639232/73.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 1; Page 387; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polynucleotide of the invention.
XX
SQ   Sequence 515 BP; 201 A; 81 C; 112 G; 118 T; 3 other;


  Query Match             18.9%;  Score 101.6;  DB 22;  Length 515;
  Best Local Similarity   61.2%;  Pred. No. 1.6e-19;
  Matches  221;  Conservative    0;  Mismatches  125;  Indels   15;  Gaps    3;

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           || | || |    | || |  ||| |||||||| ||||| |   |||||||  || | |||
Db     141 tgcagcccggtgaatctcagcaagaggaaccaccaactgacaatcaggatattgaacctg 200

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
             || ||| ||||       ||| || || |     |||  |  ||    |||| | ||
Db     201 gacaagagagagaa------ggaacacctccgatcgaagaacgtaaagtagaaggtgatt 254

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||  || ||    | ||||| || ||||| ||| ||||||  ||||||   | |
Db     255 gccaggaaatggatctggaaaagactcggagtgagcgtggagatggctctgatgtaaaag 314

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||| |   || || |||||  || | |  || ||    |||| ||| ||   |||  ||
Db     315 agaagactccacctaatcctaagcatgctaagactaaagaagcaggagatgggcagccat 374

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCA---GGCTGCTCCTATGTTGGAAATTTG- 482
           |   | ||||||||||| | ||||   |  ||   ||||| |   |  ||| |||| ||
Db     375 aagttaaaaagaagacaagctgaagctacacacatggctgatgtcacattgaaaatgtga 434

Qy     483 -----TTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
                 | |||||||| |||||||| || |      |||| | | |||||||  |||
Db     435 ctgaaaatttgaaaattctctcaataaagtttgagttttctctgaagaaaaaaaaanaaa 494

Qy     538 A 538
           |
Db     495 a 495


RESULT 13
AAH93807
ID   AAH93807 standard; cDNA; 515 BP.
XX
AC   AAH93807;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence P1001C.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
```

```
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 1; Page 385; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 515 BP; 201 A; 81 C; 112 G; 118 T; 3 other;


  Query Match             18.9%;  Score 101.6;  DB 22;  Length 515;
  Best Local Similarity   61.2%;  Pred. No. 1.6e-19;
  Matches  221;  Conservative    0;  Mismatches  125;  Indels   15;  Gaps    3;

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           || | || |    | || |  ||| ||||||| ||||| |   |||||||  || | |||
Db     141 tgcagcccggtgaatctcagcaagaggaaccaccaactgacaatcaggatattgaacctg 200

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
             || ||| ||||       ||| || || |     |||  |  ||    |||| | ||
Db     201 gacaagagagagaa------ggaacacctccgatcgaagaacgtaaagtagaaggtgatt 254

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||  || ||    | ||||| || ||||| ||| ||||||  ||||||   | |
Db     255 gccaggaaatggatctggaaaagactcggagtgagcgtggagatggctctgatgtaaaag 314

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||| |   || || |||||  || | |  || ||    |||| ||| ||  |||  ||
Db     315 agaagactccacctaatcctaagcatgctaagactaaagaagcaggagatgggcagccat 374

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCA---GGCTGCTCCTATGTTGGAAATTTG- 482
           |   | |||||||||| | ||||   |  ||   ||||| |   |  ||| |||| ||
Db     375 aagttaaaaagaagacaagctgaagctacacacatggctgatgtcacattgaaaatgtga 434

Qy     483 -----TTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
                | |||||||| |||||||| || |      |||| | | ||||||| |||
Db     435 ctgaaaatttgaaaattctctcaataaagtttgagttttctctgaagaaaaaaaaanaaa 494

Qy     538 A 538
           |
Db     495 a 495


RESULT  14
AAH85121
ID   AAH85121 standard; cDNA; 515 BP.
XX
AC   AAH85121;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human prostate-specific cDNA sequence P1001C.
XX
KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostate specific antigen; PSA; ss.
XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Retter MW,  Stolk JA,  Day CH,  Skeiky YAW,  Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 5; Page 285; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 515 BP; 201 A; 81 C; 112 G; 118 T; 3 other;


  Query Match             18.9%;  Score 101.6;  DB 22;  Length 515;
  Best Local Similarity   61.2%;  Pred. No. 1.6e-19;
  Matches  221;  Conservative    0;  Mismatches  125;  Indels   15;  Gaps    3;

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           || | || |    | || |  ||| ||||||| ||||| |   |||||||  || | |||
Db     141 tgcagcccggtgaatctcagcaagaggaaccaccaactgacaatcaggatattgaacctg 200

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
             || ||| ||||       ||| || || |     |||  |  ||    |||| | ||
Db     201 gacaagagagagaa------ggaacacctccgatcgaagaacgtaaagtagaaggtgatt 254

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||  || ||    | ||||| || ||||| ||| ||||||  ||||||   | |
Db     255 gccaggaaatggatctggaaaagactcggagtgagcgtggagatggctctgatgtaaaag 314

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||| |   || || |||||  || | |  || ||    |||| ||| ||  |||  ||
Db     315 agaagactccacctaatcctaagcatgctaagactaaagaagcaggagatgggcagccat 374

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCA---GGCTGCTCCTATGTTGGAAATTTG- 482
           |   | |||||||||| | ||||   |  ||   ||||| |   |  ||| |||| ||
Db     375 aagttaaaaagaagacaagctgaagctacacacatggctgatgtcacattgaaaatgtga 434

Qy     483 -----TTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
                | |||||||| |||||||| || |      |||| | | ||||||| |||
Db     435 ctgaaaatttgaaaattctctcaataaagtttgagttttctctgaagaaaaaaaaanaaa 494

Qy     538 A 538
           |
Db     495 a 495
```

```
RESULT 15
AAH02872
ID   AAH02872 standard; cDNA; 515 BP.
XX
AC   AAH02872;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen cDNA sequence for P1001C.
XX
KW   Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW   prostate cancer; immunogenic; cytostatic; vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200125272-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27464.
XX
PR   04-OCT-1999;   99US-0157455.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Skeiky YAW,  Reed SG,  Cheever MA;
XX
DR   WPI; 2001-245062/25.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for
PT   the treatment and diagnosis of prostate cancer -
XX
PS   Claim 4; Page 267-268; 276pp; English.
XX
CC   The present invention describes an isolated polypeptide (I) comprising
CC   at least an immunogenic portion of a prostate tumour antigen protein or
CC   its variant. (I) have cytostatic activity and can be used in vaccine
CC   production. (I), prostate tumour antigen polynucleotides, an antigen
CC   presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC   pharmaceutical composition containing (I) are useful for inhibiting the
CC   development of cancer in a patient. Antibodies specific for prostate
CC   specific proteins and oligonucleotides that hybridise to a
CC   polynucleotide that encodes a prostate specific protein are useful
CC   for detecting the presence or absence of a cancer or monitoring the
CC   progression the progression of a cancer, especially prostate cancer.
CC   AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 515 BP; 201 A; 81 C; 112 G; 118 T; 3 other;


  Query Match             18.9%;  Score 101.6;  DB 22;  Length 515;
  Best Local Similarity   61.2%;  Pred. No. 1.6e-19;
  Matches  221;  Conservative    0;  Mismatches  125;  Indels   15;  Gaps    3;

Qy    187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
          || | || |    | || |  ||| ||||||| ||||| |   |||||||  || | |||
Db    141 tgcagcccggtgaatctcagcaagaggaaccaccaactgacaatcaggatattgaacctg 200

Qy    247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
            || ||| ||||       ||| || || |     |||  |  ||    |||| | ||
Db    201 gacaagagagagaa------ggaacacctccgatcgaagaacgtaaagtagaaggtgatt 254

Qy    307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
          ||||||||  || ||    | ||||| || |||| ||| ||||||  ||||||    | |
Db    255 gccaggaaatggatctggaaaagactcggagtgagcgtggagatggctctgatgtaaaag 314

Qy    367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
          ||| |   || || |||||  || | |  || ||    |||| ||| ||   |||  ||
Db    315 agaagactccacctaatcctaagcatgctaagactaaagaagcaggagatgggcagccat 374

Qy    427 AGTGTTAAAAGAAGACACGTTGAAATGATGCA---GGCTGCTCCTATGTTGGAAATTTG- 482
          |   | ||||||||||| | ||||   |  ||   ||||| |   |  ||| |||| ||
Db    375 aagttaaaaagaagacaagctgaagctacacacatggctgatgtcacattgaaaatgtga 434

Qy    483 -----TTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
                  | ||||||||| |||||||| || |      |||| | | ||||||| |||
Db    435 ctgaaaatttgaaaattctctcaataaagtttgagttttctctgaagaaaaaaaaanaaa 494

Qy    538 A 538
          |
Db    495 a 495


Search completed: July  1, 2002, 07:11:23
Job time: 1919 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:33:34 ; Search time 14.3 Seconds
(without alignments)
779.466 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 49123

Minimum DB seq length: 0
Maximum DB seq length: 116

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	77.5	12.0	82	1	QFBO	micro glutamic aci
2	74.5	11.5	91	2	S55633	myristylated tegum
3	71	11.0	76	2	C38355	basic proline-rich
4	70	10.8	101	2	B72638	hypothetical prote
5	69.5	10.8	99	2	A55819	nonhistone chromos
6	67.5	10.4	98	2	T02437	hypothetical prote
7	66.5	10.3	79	2	A43732	BR6 protein - midg
8	66.5	10.3	102	2	A60674	insulin-like growt
9	65	10.1	72	2	S48790	troponin T, cardia
10	62.5	9.7	115	1	FAUTPC	procyclic acidic r
11	62	9.6	111	2	S15073	prothymosin alpha
12	62	9.6	112	1	TNRTA	prothymosin alpha
13	62	9.6	116	2	D72560	hypothetical prote
14	61.5	9.5	102	2	B31512	parathymosin - rat
15	61	9.4	114	2	A30229	procyclic acidic r
16	60.5	9.4	80	2	S54845	GTP cyclohydrolase
17	60.5	9.4	83	2	S78009	aggrecan - pig (fr
18	60.5	9.4	102	2	A32264	parathymosin - hum
19	59.5	9.2	96	2	E70811	hypothetical prote
20	59.5	9.2	109	1	TNBOA1	prothymosin alpha
21	59.5	9.2	110	1	TNHUA	prothymosin alpha
22	59.5	9.2	113	2	T15547	hypothetical prote
23	59.5	9.2	114	1	R5HS2H	ribosomal protein
24	59.5	9.2	114	2	F84266	50S ribosomal prot
25	59	9.1	80	2	PQ0477	pistil extensin-li
26	59	9.1	81	2	A05231	Balbiani ring b ch
27	59	9.1	104	2	C72637	hypothetical prote
28	58.5	9.1	89	2	I54187	small proline-rich
29	58.5	9.1	97	2	S11755	hypothetical prote
30	58.5	9.1	101	2	C90564	lipoprotein [impor
31	58.5	9.1	109	2	JC2232	high mobility grou
32	58	9.0	87	2	T22131	hypothetical prote
33	58	9.0	100	2	A33310	nonhistone chromos
34	57.5	8.9	106	2	S07520	host specificity p
35	57	8.8	61	2	S03097	aerC protein - Aer
36	57	8.8	74	2	S77141	hypothetical prote
37	57	8.8	94	2	AE3247	hypothetical prote
38	57	8.8	97	2	S04554	nodulin - alfalfa
39	57	8.8	109	2	T35249	hypothetical prote
40	57	8.8	111	1	VHBPP4	capsid assembly pr
41	57	8.8	111	2	A64563	hypothetical prote
42	56.5	8.7	99	2	T04512	hypothetical prote
43	56.5	8.7	102	1	HSHUP2	sperm histone P2 p
44	56.5	8.7	103	2	S08180	nonhistone chromos
45	56	8.7	59	2	I38926	von Hippel Lindau

ALIGNMENTS

RESULT 1
QFBO
micro glutamic acid-rich protein - bovine
N;Alternate names: neurofilament triplet L protein (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 22-Nov-1996
C;Accession: A02964
R;Isobe, T.; Okuyama, T.
FEBS Lett. 182, 389-392, 1985
A;Title: Brain micro glutamic acid-rich protein is the C-terminal endpiece of the
A;Reference number: A02964; MUID:85154567
A;Accession: A02964
A;Molecule type: protein
A;Residues: 1-82 <ISO>
C;Comment: This acidic protein, isolated from the cytosolic fraction of brain tis
C;Comment: The similarity of this sequence to part of the neurofilament triplet L
f the L protein.
C;Comment: Although there are, as yet, no known biological functions ascribed to
to cytoplasmic neuronal transport occurring in the presence of calcium ion and p
C;Superfamily: cytoskeletal keratin
C;Keywords: brain; coiled coil; cytosol; intermediate filament
F;1-82/Domain: neurofilament triplet L protein tail subdomain b (fragment) <SIG>

Query Match 12.0%; Score 77.5; DB 1; Length 82;
Best Local Similarity 28.9%; Pred. No. 17;
Matches 24; Conservative 15; Mismatches 27; Indels 17; Gaps 4;

```
Qy      36 EPATPEEG---EPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPD 92
           ||  | ||   |   ::::  |  | |:|||   :    :  ::|        : |:| :
Db       2 EP--PSEGEAEEEEKEKEEAEAEAEAEEEGAQEEEAEAAKEDAEE-------AKEEEGGE 52

Qy      93 GQEMDPPNPEEVKTPEEGEKQSQ 115
           |:|      || |  || ||: :
Db      53 GEE-----AEETKEAEEEEKKDE 70
```

RESULT 2
S55633
myristylated tegument protein 38 - equine herpesvirus 2
C;Species: equine herpesvirus 2
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C;Accession: S55633
R;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.
J. Mol. Biol. 249, 520-528, 1995
A;Title: The DNA sequence of equine herpesvirus 2.
A;Reference number: S55594; MUID:95302501
A;Accession: S55633
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-91 <TEL>
A;Cross-references: GB:U20824; NID:g695172; PIDN:AAC13826.1; PID:g695211

```
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995


  Query Match             11.5%;  Score 74.5;  DB 2;  Length 91;
  Best Local Similarity   30.8%;  Pred. No. 30;
  Matches   24;  Conservative    9;  Mismatches   38;  Indels    7;  Gaps    2;

Qy      36 EPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQE 95
           :| |  ||:|   |::    :|| ||   |     |   : |        | ::| || |
Db      13 DPLTDVEGQPINVREEFEMFEEG-DEATEACVFLNPNMGTDEYD------EEDEGGDGGE 65

Qy      96 MDPPNPEEVKTPEEGEKQ 113
             | ||   ||    |:
Db      66 GREPQPEVKTTPYPKRKK 83


RESULT   3
C38355
basic proline-rich peptide II-2 - human
C;Species: Homo sapiens (man)
C;Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 12-Apr-1995
C;Accession: C38355
R;Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.
Biochemistry 30, 3351-3356, 1991
A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov
A;Reference number: A38355; MUID:91190884
A;Accession: C38355
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-76 <KAU>
C;Superfamily: proline-rich protein


  Query Match             11.0%;  Score 71;  DB 2;  Length 76;
  Best Local Similarity   27.4%;  Pred. No. 47;
  Matches   20;  Conservative    9;  Mismatches   34;  Indels   10;  Gaps    2;

Qy      44 EPATQRQDPAA----AQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPP 99
           |  :| : |:      |    :| :  ||| |    : || |  |     ||      ||
Db       5 EDVSQEESPSLIAGNPQGPSPQGGNKPQGPPPPPPGKPQGPPPQGGNKPQGP------PP 58

Qy     100 NPEEVKTPEEGEK 112
             :    | :|:|
Db      59 PGKPQGPPPQGDK 71


RESULT   4
B72638
hypothetical protein APE0539 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: B72638
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339
A;Accession: B72638
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-101 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79506.1; PID:d1043292; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0539


  Query Match             10.8%;  Score 70;  DB 2;  Length 101;
  Best Local Similarity   25.8%;  Pred. No. 72;
  Matches   17;  Conservative   16;  Mismatches   31;  Indels    2;  Gaps    2;

Qy      49 RQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPE 108
           : :|    |||| |    :| :|   :|: |  : | |   ||  : : |    |::  :

Db       5 KHEPNPQGEGED-GPEGARGEEPGHPAQDDGEARRGAEAA-GPGHEALLPRQVREMRGAD 62

Qy     109 EGEKQS 114
           : |:::
Db      63 QPEEEA 68


RESULT   5
A55819
nonhistone chromosomal protein CHMG-I - midge (Chironomus tentans)
N;Alternate names: high mobility group protein I/Y homolog
C;Species: Chironomus tentans
C;Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 21-Jul-2000
C;Accession: A55819; S47253; S47252
R;Claus, P.; Schulze, E.; Wisniewski, J.R.
J. Biol. Chem. 269, 33042-33048, 1994
A;Title: Insect proteins homologous to mammalian high mobility group proteins I,
A;Reference number: A55819; MUID:95105193
A;Accession: A55819
A;Molecule type: mRNA
A;Residues: 1-99 <CLA>
A;Cross-references: EMBL:Z36898; NID:g534932; PIDN:CAA85365.1; PID:g534933
A;Experimental source: clone pCW126
R;Claus, P.; Schulze, E.; Wisniewski, J.R.
submitted to the EMBL Data Library, August 1994
A;Description: Insect proteins homologous to mammalian high mobility group prote
A;Reference number: S47252
A;Accession: S47253
A;Molecule type: DNA
A;Residues: 1-91,'A',93-99 <CL2>
A;Cross-references: EMBL:Z36897; NID:g534886; PID:g534887
A;Experimental source: clone pCWG106
C;Genetics:
A;Introns: 15/3; 76/1
A;Note: suggested by Southern blot analysis in Ref A55819 to be single copy gen
C;Keywords: chromosomal protein; DNA binding
F;8-13/Region: DNA-binding motif (K/R-G-R-G-R-P)
F;55-60/Region: DNA-binding motif (K/R-G-R-G-R-P)
F;75-80/Region: DNA-binding motif (K/R-G-R-G-R-P)


  Query Match             10.8%;  Score 69.5;  DB 2;  Length 99;
  Best Local Similarity   24.7%;  Pred. No. 77;
  Matches   22;  Conservative   21;  Mismatches   43;  Indels    3;  Gaps    3;

Qy      10 RPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQ-G 68
           | || :  :| |   |   ::   :||   |:: | :|: :: || ::|   :   :
Db      10 RGRPAK-AKPEETASPKAAKKEEKKVE-EVPKKIEESTKPENGAAPKKGRGRPSKGDKAA 67

Qy      69 PKPEAHSQEQGHPQTGCECEDGPDGQEMD 97
           ||  |  : :| |    :  |  | :|::
Db      68 PKRPASGKGRGRPAKNAKKVDDADTEEVN 96


RESULT   6
T02437
hypothetical protein At2g41420 [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein T26J13.1
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02437; E84841
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.
submitted to the EMBL Data Library, June 1998
A;Description: Arabidopsis thaliana chromosome II BAC T26J13 genomic sequence.
A;Reference number: Z14673
A;Accession: T02437
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-98 <ROU>
A;Cross-references: EMBL:AC004625; NID:g3241939; PID:g3241940
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Ta
```

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: E84841
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-98 <STO>
A;Cross-references: GB:AE002093; NID:g3241940; PIDN:AAC23727.1; GSPDB:GN00139
C;Genetics:
A;Gene: T26J13.1; At2g41420
A;Map position: 2
A;Introns: 16/1; 84/2

Query Match 10.4%; Score 67.5; DB 2; Length 98;
Best Local Similarity 29.1%; Pred. No. 1.1e+02;
Matches 23; Conservative 10; Mismatches 33; Indels 13; Gaps 3;

```
Qy      40 PEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQE----QGHPQTGCECEDGPDGQE 95
           |  | |  |   |    ||  : |   ||  |: : |:    ||:|| |   :  |
Db       7 PPVGVPPPQGYPP----EGYPKDAYPPQGYPPQGYPQQGYPPQGYPQQGYPQQGYP---- 58

Qy      96 MDPPNPEEVKTPEEGEKQS 114
             |  |:    |:  ::||
Db      59 -PPYAPQYPPPPQHQQQQS 76
```

RESULT 7
A43732
BR6 protein - midge (Chironomus tentans) (fragments)
C;Species: Chironomus tentans
C;Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 17-Mar-1999
C;Accession: A43732
R;Lendahl, U.; Wieslander, L.
Cell 36, 1027-1034, 1984
A;Title: Balbiani ring 6 gene in Chironomus tentans: a diverged member of the Balbiani r
A;Reference number: A43732; MUID:84156521
A;Accession: A43732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-79 <LEN>
A;Cross-references: EMBL:K01692

Query Match 10.3%; Score 66.5; DB 2; Length 79;
Best Local Similarity 28.9%; Pred. No. 1e+02;
Matches 22; Conservative 9; Mismatches 32; Indels 13; Gaps 3;

```
Qy      12 RPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKP 71
           || |   | |   | |||:       |  ||| |    |::|    : :||     :  |
Db      12 RPERPERPEEPERPERPER------PERPEEPE----REEPEREPKCDDE---MREKVKR 58

Qy      72 EAHSQEQGHPQTGCEC 87
             :: :      |||
Db      59 RCDNENRRFDARRCEC 74
```

RESULT 8
A60674
insulin-like growth factor I receptor - mouse (fragment)
N;Alternate names: IGF-1 receptor
C;Species: Mus musculus (house mouse)
C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 23-May-1997
C;Accession: A60674
R;Telford, N.A.; Hogan, A.; Franz, C.R.; Schultz, G.A.
Mol. Reprod. Dev. 27, 81-92, 1990
A;Title: Expression of genes for insulin and insulin-like growth factors and receptors i
A;Reference number: A60674; MUID:91064061
A;Accession: A60674
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-102 <TEL>
C;Superfamily: insulin receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; membrane protein

Query Match 10.3%; Score 66.5; DB 2; Length 102;
Best Local Similarity 28.4%; Pred. No. 1.3e+02;
Matches 27; Conservative 13; Mismatches 30; Indels 25; Gaps 6;

```
Qy       9 YRPRPRRYVEPP--EMIGPMRPEQ----------FSDEVEPATPEEGEPATQRQDPAAAQ 56
           | |:    : |   |:|| :: |           :|:| :|  ||| |   :  :|
Db       2 YNPK----IRPSFLEIIGSIKDEMEPSFQEVSFYYSEENKPPEPEELEMELE-MEPENIG 56

Qy      57 EGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGP 91
           |   :  |    | || ||   ||     : |:||
Db      57 ELPLDPRSQSSLPLPERHS---GH-----KAENGP 83
```

RESULT 9
S48790
troponin T, cardiac muscle (clone HTNT3) - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Aug-1999
C;Accession: S48790
R;Farza, H.
submitted to the EMBL Data Library, June 1994
A;Reference number: S48790
A;Accession: S48790
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-72 <FAR>
A;Cross-references: EMBL:X79858; NID:g587429; PIDN:CAA56238.1; PID:g587430
C;Superfamily: troponin T
C;Keywords: cardiac muscle; heart

Query Match 10.1%; Score 65; DB 2; Length 72;
Best Local Similarity 34.1%; Pred. No. 1.2e+02;
Matches 14; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

```
Qy      33 DEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEA 73
           || | |  |: |   : ::  |  | |:|   |  ||  |:
Db      31 DEQEEAAEEDAEAEAETEETRAEDEEEEEAKEAEDGPMEES 71
```

RESULT 10
FAUTPC
procyclic acidic repetitive protein precursor (clone pAP3) - Trypanosoma brucei b
N;Alternate names: procyclin
C;Species: Trypanosoma brucei brucei
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 22-Jun-1999
C;Accession: S14896; S06170
R;Koenig, E.; Delius, H.; Carrington, M.; Williams, R.O.; Roditi, I.
Nucleic Acids Res. 17, 8727-8739, 1989
A;Title: Duplication and transcription of procyclin genes in Trypanosoma brucei.
A;Reference number: S14896; MUID:90067841
A;Accession: S14896
A;Molecule type: DNA
A;Residues: 1-115 <KOE>
A;Cross-references: EMBL:X16015; NID:g10512; PIDN:CAA34147.1; PID:g10513
C;Genetics:
A;Gene: ProA
C;Superfamily: procyclic acidic repetitive protein
C;Keywords: glycoprotein; tandem repeat
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-115/Product: procyclic acidic repetitive protein #status predicted <MAT>
F;63-92/Region: 2-residue repeats (E-P)
F;56/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 9.7%; Score 62.5; DB 1; Length 115;
Best Local Similarity 28.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 7; Mismatches 24; Indels 23; Gaps 3;

```
Qy      53 AAAQEG-EDEGASAG----------------QGPKPEAHSQEQGHPQTGCECEDGPDGQE 95
           ||| || ||:| : |                 |  |:   : :  |:   | |  |:
Db      25 AAAAEGPEDKGLTKGGKGKGEKGTKVGADDTNGTDPDPEPEPEPEPEPEPEPEPEPE--- 81

Qy      96 MDPPNPEEVKTPEEG 110
              | ||    || |
Db      82 ---PEPEPEPEPEPG 93
```

RESULT 11
S15073
prothymosin alpha - mouse
C;Species: Mus musculus (house mouse)
C;Date: 18-Feb-1994 #sequence_revision 02-Dec-1994 #text_change 16-Jul-1999
C;Accession: S15073; S12883; S06666
R;Schmidt, G.; Werner, D.
Biochim. Biophys. Acta 1088, 442-444, 1991
A;Title: Nucleotide sequence of the murine prothymosin alpha cDNA and its deduced primar
A;Reference number: S15073; MUID:91198154
A;Accession: S15073
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-111 <SCH>
A;Cross-references: EMBL:X56135; NID:g54796; PIDN:CAA39601.1; PID:g54797
R;Low, T.L.K.; Pan, T.L.; Lin, Y.S.
FEBS Lett. 273, 1-5, 1990
A;Title: Depression of prothymosin-alpha production in murine thymus correlates with sta
A;Reference number: S12883; MUID:91032153
A;Accession: S12883
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-106,'K',107-111 <LOW>
R;Makarova, T.; Grebenshikov, N.; Egorov, C.; Vartapetian, A.; Bogdanov, A.
FEBS Lett. 257, 247-250, 1989
A;Title: Prothymosin alpha is an evolutionary conserved protein covalently linked to a s
A;Reference number: S06666; MUID:90060328
A;Accession: S06666
A;Molecule type: protein
A;Residues: 16-19;23-68;91-100,'E',102-103 <MAK>
C;Superfamily: prothymosin alpha
C;Keywords: nucleus; phosphoprotein

  Query Match   9.6%; Score 62; DB 2; Length 111;
  Best Local Similarity   24.7%; Pred. No. 3.1e+02;
  Matches   19; Conservative   13; Mismatches   43; Indels   2; Gaps   1;

```
Qy      35 VEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQ 94
           || |  | |  |    :    : || |  :     : |   :|:   :   | ||| : :
Db      24 VEEA--ENGRDAPANGNAQNEENGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDE 81

Qy      95 EMDPPNPEEVKTPEEGE 111
           | : |  : |   :| :
Db      82 EAEAPTGKRVAEDDEDD 98
```

RESULT 12
TNRTA
prothymosin alpha - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1987 #sequence_revision 02-Dec-1994 #text_change 16-Jun-2000
C;Accession: A32265; A37401; I53187; A01519; A20738; S13014
R;Frangou-Lazaridis, M.; Clinton, M.; Goodall, G.J.; Horecker, B.L.
Arch. Biochem. Biophys. 263, 305-310, 1988
A;Title: Prothymosin alpha-and parathymosin: amino acid sequences deduced from the clone
A;Reference number: A32265; MUID:88239673
A;Accession: A32265
A;Molecule type: mRNA
A;Residues: 1-112 <FRA>
A;Cross-references: GB:M33024; NID:g206347; PIDN:AAA41931.1; PID:g206348
R;Oikawa, M.; Dargan, C.; Ny, T.; Hsueh, A.J.W.
Endocrinology 127, 2350-2356, 1990
A;Title: Expression of gonadotropin-releasing hormone and prothymosin-alpha messenger ri
A;Reference number: A37401; MUID:91031267
A;Accession: A37401
A;Molecule type: mRNA
A;Residues: 1-63,'K',65-67,'K',69-83 <OIK>
A;Cross-references: GB:M60664
A;Note: 64-Glu and 68-Glu were also found
R;Eilers, M.; Schirm, S.; Bishop, J.M.
EMBO J. 10, 133-141, 1991
A;Title: The MYC protein activates transcription of the alpha-prothymosin gene.
A;Reference number: I53187; MUID:91114696
A;Accession: I53187
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-112 <EIL>
A;Cross-references: EMBL:X55326; NID:g57661; PIDN:CAA39028.1; PID:g57662
A;Experimental source: cell line RAT1A
A;Note: in GenBank entry RRPROTHYO, release 113.0, the source is designated as R
R;Haritos, A.A.; Blacher, R.; Stein, S.; Caldarella, J.; Horecker, B.L.
Proc. Natl. Acad. Sci. U.S.A. 82, 343-346, 1985
A;Title: Primary structure of rat thymus prothymosin alpha.
A;Reference number: A01519; MUID:85113197
A;Accession: A01519
A;Molecule type: protein
A;Residues: 2-61,'G',63-64,'BGGZZZZZZZZZZN',77-112 <HAR1>
A;Note: it is possible that both Asp and Asn are present at positions 29 and 36
R;Haritos, A.A.; Goodall, G.J.; Horecker, B.L.
Proc. Natl. Acad. Sci. U.S.A. 81, 1008-1011, 1984
A;Title: Prothymosin alpha: isolation and properties of the major immunoreactive
A;Reference number: A20738; MUID:84144830
A;Accession: A20738
A;Molecule type: protein
A;Residues: 2-31 <HAR2>
R;Palvimo, J.; Linnala-Kankkunen, A.
FEBS Lett. 277, 257-260, 1990
A;Title: Identification of a low-M(r) acidic nuclear protein as prothymosin alph
A;Reference number: S13014; MUID:91099511
A;Accession: S13014
A;Molecule type: protein
A;Residues: 2-112 <PAL>
C;Comment: This highly acidic protein was originally studied as a putative thymi
arent function may not be physiological. The native protein is expressed in a br
 precise role is not yet known.
C;Superfamily: prothymosin alpha
C;Keywords: blocked amino end; nucleus; phosphoprotein
F;2-112/Product: prothymosin alpha #status experimental <MAT>
F;2-29/Product: thymosin alpha-1 #status experimental <TMA>
F;2/Modified site: blocked amino end (Ser) (in mature form) (probably acetylated

  Query Match   9.6%; Score 62; DB 1; Length 112;
  Best Local Similarity   24.7%; Pred. No. 3.1e+02;
  Matches   19; Conservative   13; Mismatches   43; Indels   2; Gaps   1;

```
Qy      35 VEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQ 94
           || |  | |  |    :    : || |  :     : |   :|:   :   | ||| : :
Db      24 VEEA--ENGRDAPANGNAQNEENGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDE 81

Qy      95 EMDPPNPEEVKTPEEGE 111
           | : |  : |   :| :
Db      82 EAEAPTGKRVAEDDEDD 98
```

RESULT 13
D72560
hypothetical protein APE1766 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: D72560
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon,
A;Reference number: A72450; MUID:99310339

A;Accession: D72560
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-116 <KAW>
A;Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80769.1; PID:d1044555; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1766

Query Match 9.6%; Score 62; DB 2; Length 116;
Best Local Similarity 24.4%; Pred. No. 3.2e+02;
Matches 22; Conservative 7; Mismatches 41; Indels 20; Gaps 2;

```
Qy      41 EEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQ-------------TGCEC 87
           ::|    :|    : : || ||   |  |: :    | |  |               |
Db       3 KQGASLRRRGARHSYRRGEPEGGDRGSQPRQDLPGCEGGGAQRIPNSQPRSTPAIARLLC 62

Qy      88 EDGPDGQEMD-------PPNPEEVKTPEEG 110
           | | ||   |       || |     |  |
Db      63 EGGEDGCRDDRRRRHSHPPEPRAGGLPARG 92
```

RESULT 14
B31512
parathymosin - rat
N;Alternate names: 11K zinc-binding protein; phosphofructokinase-1-inactivating protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Feb-1990 #sequence_revision 02-Dec-1994 #text_change 16-Jul-1999
C;Accession: S05212; B31512; A35060; B32265; A05095; A22075
R;Trompeter, H.I.; Brand, I.A.; Soeling, H.D.
FEBS Lett. 253, 63-66, 1989
A;Title: The primary sequence of the PFK-1 inactivating zinc-binding protein as deduced
A;Reference number: S05212; MUID:89338759
A;Accession: S05212
A;Molecule type: mRNA
A;Residues: 1-102 <TRO>
A;Cross-references: EMBL:X16481; NID:g55538; PIDN:CAA34501.1; PID:g55539
R;Panneerselvam, C.; Clinton, M.; Wellner, D.; Horecker, B.L.
Biochem. Biophys. Res. Commun. 155, 539-545, 1988
A;Title: Bovine parathymosin: amino acid sequence and comparison with rat parathymosin.
A;Reference number: A90147; MUID:88339930
A;Accession: B31512
A;Molecule type: protein
A;Residues: 2-102 <PAN>
R;Wellner, D.; Panneerselvam, C.; Horecker, B.L.
Proc. Natl. Acad. Sci. U.S.A. 87, 1947-1949, 1990
A;Title: Sequencing of peptides and proteins with blocked N-terminal amino acids: N-acet
A;Reference number: A35060; MUID:90175413
A;Accession: A35060
A;Molecule type: protein
A;Residues: 2-8 <WEL>
R;Frangou-Lazaridis, M.; Clinton, M.; Goodall, G.J.; Horecker, B.L.
Arch. Biochem. Biophys. 263, 305-310, 1988
A;Title: Prothymosin alpha-and parathymosin: amino acid sequences deduced from the clone
A;Reference number: A32265; MUID:88239673
A;Accession: B32265
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 23-102 <FRA>
A;Cross-references: GB:M33025; NID:g206032; PIDN:AAA41810.1; PID:g206033
R;Komiyama, T.; Pan, L.X.; Haritos, A.A.; Wideman, J.W.; Pan, Y.C.E.; Chang, M.; Rogers,
Proc. Natl. Acad. Sci. U.S.A. 83, 1242-1245, 1986
A;Reference number: A05095; MUID:86149281
A;Accession: A05095
A;Molecule type: protein
A;Residues: 2,'KSE',6-54,'GGGBBBBBBBZZZZZZZZZZZZZZZ',75-102 <KOM>
R;Haritos, A.A.; Salvin, S.B.; Blacher, R.; Stein, S.; Horecker, B.L.
Proc. Natl. Acad. Sci. U.S.A. 82, 1050-1053, 1985
A;Title: Parathymosin alpha: a peptide from rat tissues with structural homology to prot
A;Reference number: A22075; MUID:85140226
A;Accession: A22075
A;Molecule type: protein
A;Residues: 2,'KSE',6-31 <HAR>
C;Superfamily: prothymosin alpha
C;Keywords: acetylated amino end; zinc
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experiment

Query Match 9.5%; Score 61.5; DB 2; Length 102;
Best Local Similarity 25.7%; Pred. No. 3.1e+02;
Matches 19; Conservative 15; Mismatches 23; Indels 17; Gaps 4;

```
Qy      41 EEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGP----DGQEM 96
           || |   : ::   |::|||:     :| : :   :|        | ::||      :|
Db      39 EEEENGAEEEEEETAEDGEDD----DEGDEEDEEEEE--------EEDEGPVRKRTAEEE 86

Qy      97 DPPNPEEVKTPEEG 110
           |  :|:  || | |
Db      87 DEADPKRQKT-ENG 99
```

RESULT 15
A30229
procyclic acidic repetitive protein A-alpha - Trypanosoma brucei
C;Species: Trypanosoma brucei
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 13-Aug-1999
C;Accession: A30229; S21539
R;Mowatt, M.R.; Wisdom, G.S.; Clayton, C.E.
Mol. Cell. Biol. 9, 1332-1335, 1989
A;Title: Variation of tandem repeats in the developmentally regulated procyclic a
A;Reference number: A30229; MUID:89261740
A;Accession: A30229
A;Molecule type: DNA
A;Residues: 1-114 <MOW>
A;Cross-references: GB:M25787; NID:g341399; PIDN:AAA53283.1; PID:g576546
R;Vijayasarathy, S.; Ernest, I.; Itzhaki, J.; Sherman, D.; Mowatt, M.R.; Michels,
submitted to the EMBL Data Library, April 1990
A;Description: The genes encoding fructose bisphosphate aldolase in trypanosoma b
A;Reference number: S21538
A;Accession: S21539
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-114 <VIJ>
A;Cross-references: EMBL:X52584; NID:g10480; PIDN:CAA36814.1; PID:g10481
C;Superfamily: procyclic acidic repetitive protein

Query Match 9.4%; Score 61; DB 2; Length 114;
Best Local Similarity 34.4%; Pred. No. 3.7e+02;
Matches 22; Conservative 8; Mismatches 24; Indels 10; Gaps 4;

```
Qy      60 DEGAS-----AGQGPKPE---AHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGE 111
           || ||      |:| : |      :| |  :|| | | ||:    :   |||   ||| |
Db      29 DESASNVIVKGGKGKEREDGPEEPEETGPEETGPE-ETGPEETGPEETGPEET-GPEETE 86

Qy     112 KQSQ 115
             : :
Db      87 PEPE 90
```

Search completed: July 1, 2002, 06:37:04
Job time: 210 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:36:39 ; Search time 12.61 Seconds
(without alignments)
356.183 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 16058

Minimum DB seq length: 0
Maximum DB seq length: 116

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	646	100.0	116	1	GGE2_HUMAN	Q13066	homo sapien
2	120	18.6	102	1	GGC1_HUMAN	O60829	homo sapien
3	72.5	11.2	85	1	MRKD_RAT	P20468	rattus norv
4	70.5	10.9	96	1	PRP5_HUMAN	P04281	homo sapien
5	62.5	9.7	61	1	PRPE_HUMAN	P02811	homo sapien
6	62.5	9.7	115	1	PARX_TRYBB	P14043	trypanosoma
7	62	9.6	110	1	THYA_MOUSE	P26350	mus musculu
8	62	9.6	111	1	THYA_RAT	P06302	rattus norv
9	61.5	9.5	101	1	THYP_RAT	P04550	rattus norv
10	61	9.4	113	1	MEA1_PIG	Q95313	sus scrofa
11	61	9.4	114	1	PARA_TRYBB	P18764	trypanosoma
12	60.5	9.4	83	1	PGCA_PIG	Q29011	sus scrofa
13	60.5	9.4	101	1	THYP_HUMAN	P20962	homo sapien
14	59.5	9.2	109	1	THYA_BOVIN	P01252	bos taurus
15	59.5	9.2	114	1	RL12_HALHA	P05768	halobacteri
16	58.5	9.1	89	1	CORA_HUMAN	P35321	homo sapien
17	58.5	9.1	109	1	HMGC_HUMAN	P52926	homo sapien
18	58	9.0	99	1	HG14_HUMAN	P05114	homo sapien
19	57.5	8.9	101	1	THYP_BOVIN	P08814	bos taurus
20	57.5	8.9	106	1	VHSB_BPT3	P20322	bacteriopha
21	57	8.8	61	1	AERC_AERSO	P09165	aeromonas s
22	57	8.8	97	1	NO75_MEDSA	P11728	medicago sa
23	57	8.8	111	1	VG31_BPT4	P17313	bacteriopha
24	56.5	8.7	102	1	HG14_CHICK	P12274	gallus gall
25	56.5	8.7	102	1	HSP2_HUMAN	P04554	homo sapien
26	56.5	8.7	110	1	THYA_HUMAN	P06454	homo sapien
27	56	8.7	114	1	ET3_RABIT	P19998	oryctolagus
28	55.5	8.6	102	1	DAP1_HUMAN	P51397	homo sapien
29	55.5	8.6	102	1	HSP2_MACMU	P35297	macaca mula
30	55.5	8.6	103	1	HSP2_MACNE	P35298	macaca neme
31	55	8.5	82	1	MT21_ORYSA	P94029	oryza sativ
32	54.5	8.4	89	1	CORB_HUMAN	P22528	homo sapien
33	54.5	8.4	89	1	CORN_MACMU	P35322	macaca mula
34	54.5	8.4	95	1	HMGY_MOUSE	P17095	mus musculu
35	54.5	8.4	100	1	HG14_BOVIN	P02316	bos taurus
36	54	8.4	104	1	HG15_CHICK	P12902	gallus gall
37	53.5	8.3	84	1	VE4_HPV1A	P06923	human papil
38	53.5	8.3	86	1	CA1C_BOVIN	P25508	bos taurus
39	53.5	8.3	104	1	LE11_HELAN	P46515	helianthus
40	53.5	8.3	107	1	DPD4_HUMAN	Q9hcu8	homo sapien
41	53	8.2	97	1	CORN_PIG	P35323	sus scrofa
42	53	8.2	105	1	COLL_HSVS7	P25050	herpesvirus
43	52.5	8.1	66	1	ITH2_HIRME	P09944	hirudo medi
44	52.5	8.1	110	1	DCD_HUMAN	P81605	homo sapien
45	52	8.0	65	1	WH11_CANAL	P43074	candida alb

ALIGNMENTS

RESULT 1
GGE2_HUMAN
ID GGE2_HUMAN STANDARD; PRT; 116 AA.
AC Q13066;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-2 protein (G antigen 2).
GN GAGE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC T LYMPHOCYTES.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U19143; AAA82745.1; -.
DR MIM; 604244; -.
KW Multigene family; Antigen.
SQ SEQUENCE 116 AA; 12786 MW; DD305D5CA29AF19A CRC64;

Query Match 100.0%; Score 646; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 3.8e-38;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60

Qy     61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
```

RESULT 2

```
GGC1_HUMAN
ID   GGC1_HUMAN     STANDARD;      PRT;   102 AA.
AC   O60829;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G antigen family C 1 protein (Prostate-associated gene protein 4)
DE   (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN   GAGEC1 OR PAGE4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98393718; PubMed=9724777;
RA   Brinkmann U., Vasmatzis G., Lee B., Yerushalmi N., Essand M.,
RA   Pastan I.;
RT   "PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
RT   normal and neoplastic prostate, testis, and uterus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RA   Meindl A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
CC       FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
CC       UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CC       CANCER, AND UTERINE CANCER.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF275258; AAF88037.1; -.
DR   EMBL; AJ005894; CAA06751.1; -.
DR   EMBL; AF238380; AAF62541.1; -.
DR   MIM; 300287; -.
SQ   SEQUENCE   102 AA;  11153 MW;  CE5D07AFBF73301B CRC64;

  Query Match             18.6%;  Score 120;  DB 1;  Length 102;
  Best Local Similarity   34.5%;  Pred. No. 0.02;
  Matches   39;  Conservative   10;  Mismatches   52;  Indels   12;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           || | ||  | |     | |:::          | |   :: || |  ||    |  |
Db       1 MSARVRSRSRGRGDGQ-EAPDVVA---------FVAPGESQQEEPPTDNQDIEPGQ--ER 48

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           ||    :  | |   ||    :|  |  || | :|  ||||:  || | |: |
Db      49 EGTPPIEERKVEGDCQEMDLEKTRSERGDGSDVKEKTPPNPKHAKTKEAGDGQ 101


RESULT   3
MRKD_RAT
ID   MRKD_RAT       STANDARD;      PRT;    85 AA.
AC   P20668;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-FEB-1991 (Rel. 17, Last annotation update)
DE   Protein kinase C substrate 80 kDa protein (Fragments).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=90005952; PubMed=2676596;
RA   Erusalimsky J.D., Morris C., Perks K., Brown R., Brooks S.,
RA   Rozengurt E.;
RT   "Internal amino acid sequence analysis of the 80 kDa protein kinase C
RT   substrate from rat brain: relationship to the 87 kDa substrate from
RT   bovine brain.";
RL   FEBS Lett. 255:149-153(1989).
CC   -!- FUNCTION: CELLULAR SUBSTRATE FOR PROTEIN KINASE C.
CC   -!- SIMILARITY: TO 87 MYRISTOYLATED ALANINE-RICH C KINASE SUBSTRATE
CC       (MARCKS).
DR   PIR; S05681; S05681.
KW   Phosphorylation.
FT   NON_TER       1      1
FT   NON_CONS     20     21
FT   NON_CONS     49     50
FT   NON_CONS     73     74
SQ   SEQUENCE   85 AA;  7593 MW;  39501D788C12E268 CRC64;

  Query Match             11.2%;  Score 72.5;  DB 1;  Length 85;
  Best Local Similarity   27.4%;  Pred. No. 29;
  Matches   23;  Conservative    8;  Mismatches   38;  Indels   15;  Gaps    3;

Qy      36 EPATPEEGEPATQRQ----------DPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGC 85
             | |    || |                 |||  ||  | :  :| : |  ::|   |  |
Db       2 EAAAERPGEAAVASSPGXAAAAAAGGDAAAAPGEQAGGAGAEGAEGEEKAEE---PAPGA 58

Qy      86 ECEDGPD--GQEMDPPNPEEVKTP 107
             :| |   | | :  :||    |
Db      59 TADDAPSAAGPEQEAXSPEAPPAP 82


RESULT   4
PRP5_HUMAN
ID   PRP5_HUMAN     STANDARD;      PRT;    96 AA.
AC   P04281;
DT   20-MAR-1987 (Rel. 04, Created)
DT   20-MAR-1987 (Rel. 04, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   Basic proline-rich peptide IB-1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=86243355; PubMed=3521730;
RA   Kauffman D., Hofmann T., Bennick A., Keller P.;
RT   "Basic proline-rich proteins from human parotid saliva: complete
RT   covalent structures of proteins IB-1 and IB-6.";
RL   Biochemistry 25:2387-2392(1986).
CC   -!- DOMAIN: CONTAINS 21-RESIDUE REPEATS, TWO OF WHICH HAVE INTERNAL
CC       7-RESIDUE REPEATS.
CC   -!- MISCELLANEOUS: THERE ARE NINE BASIC PROLINE-RICH PEPTIDES ISOLATED
CC       FROM THE SALIVA; THIS PEPTIDE IS DESIGNATED IB-1.
DR   PIR; A03293; PIHUB1.
KW   Phosphorylation; Repeat; Parotid gland; Saliva.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       8      8       PHOSPHORYLATION.
FT   REPEAT       15     35       APPROXIMATE.
FT   REPEAT       36     56       APPROXIMATE.
FT   REPEAT       57     77       APPROXIMATE.
SQ   SEQUENCE   96 AA;  9530 MW;  FF3B33F68DE104F8 CRC64;

  Query Match             10.9%;  Score 70.5;  DB 1;  Length 96;
  Best Local Similarity   27.7%;  Pred. No. 43;
  Matches   26;  Conservative    8;  Mismatches   43;  Indels   17;  Gaps    4;

Qy      29 EQFSDEVEP---------ATPEEG-EPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQ 78
```

```
          |  | |  |         | |: | :|      |   |    :| :  ||| |    : |
Db       5 EDVSQEESPSLIAGNPQGAPPQGGNKPQGPPSPPGKPQGPPPQGGNQPQGPPPPP-GKPQ 63

Qy      79 GHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEK 112
           | |  |     ||      ||  :    | :|:|
Db      64 GPPPQGGNKPQGP------PPPGKPQGPPPQGDK 91


RESULT   5
PRPE_HUMAN
ID   PRPE_HUMAN     STANDARD;      PRT;    61 AA.
AC   P02811;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Basic proline-rich peptide P-E (IB-9).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Saliva;
RX   MEDLINE=83101329; PubMed=6924859;
RA   Kauffman D., Wong R., Bennick A., Keller P.;
RT   "Basic proline-rich proteins from human parotid saliva: complete
RT   covalent structure of protein IB-9 and partial structure of protein
RT   IB-6, members of a polymorphic pair.";
RL   Biochemistry 21:6558-6562(1982).
RN   [2]
RP   SEQUENCE.
RC   TISSUE=Saliva;
RX   MEDLINE=83007119; PubMed=7118863;
RA   Isemura S., Saitoh E., Sanada K.;
RT   "Fractionation and characterization of basic proline-rich peptides of
RT   human parotid saliva and the amino acid sequence of proline-rich
RT   peptide P-E.";
RL   J. Biochem. 91:2067-2075(1982).
CC   -!- MISCELLANEOUS: PEPTIDES IB-9 AND P-E ARE THE SAME PEPTIDE.
DR   PIR; A03292; PIHUB9.
KW   Repeat; Parotid gland; Saliva.
FT   REPEAT        1     21       APPROXIMATE.
FT   REPEAT       22     42       APPROXIMATE.
FT   REPEAT       43     61       APPROXIMATE.
SQ   SEQUENCE   61 AA;  6024 MW;  F883F78617E05327 CRC64;


  Query Match             9.7%;  Score 62.5;  DB 1;  Length 61;
  Best Local Similarity   23.6%;  Pred. No. 1e+02;
  Matches   21;  Conservative   12;  Mismatches   25;  Indels   31;  Gaps    4;

Qy      19 PPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQ 78
           | :  ||  | |  :: :   |  |:|                       ||| |:  :: |
Db       3 PGKPQGP--PPQGGNQPQGPPPPPGKP--------------------QGPPPQGGNRPQ 39

Qy      79 GHPQTGCECEDGPDGQEMDPPNPEEVKTP 107
           | |  |      | |    ||  :: ::|
Db      40 GPPPPG-----KPQG---PPPQGDKSRSP 60


RESULT   6
PARX_TRYBB
ID   PARX_TRYBB     STANDARD;      PRT;   115 AA.
AC   P14043;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Procyclic form specific polypeptide precursor (Procyclin) (PARP).
GN   PROA.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=227(ILTAT1);
RX   MEDLINE=90067841; PubMed=2573878;
RA   Koenig E., Delius H., Carrington M., Williams R.O., Roditi I.;
RT   "Duplication and transcription of procyclin genes in Trypanosoma
RT   brucei.";
RL   Nucleic Acids Res. 17:8727-8739(1989).
CC   -!- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC       DIFFERENTIATION IN THE INSECT VECTOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16015; CAA34147.1; -.
DR   PIR; S14896; FAUTPC.
KW   Signal; Antigen; Repeat; Glycoprotein; GPI-anchor.
FT   SIGNAL        1     27
FT   CHAIN        28     93       PROCYCLIC FORM SPECIFIC POLYPEPTIDE.
FT   PROPEP       94    115
FT   DOMAIN       59     92       17 X 2 AA TANDEM REPEATS OF [DE]-P.
FT   CARBOHYD     56     56       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   115 AA;  11714 MW;  6D39F0731CF0509E CRC64;


  Query Match             9.7%;  Score 62.5;  DB 1;  Length 115;
  Best Local Similarity   28.0%;  Pred. No. 1.8e+02;
  Matches   21;  Conservative    7;  Mismatches   24;  Indels   23;  Gaps    3;

Qy      53 AAAQEG-EDEGASAG----------------QGPKPEAHSQEQGHPQTGCECEDGPDGQE 95
           ||| || ||:| : |                |  |:   : :  |:   | |  |:
Db      25 AAAAEGPEDKGLTKGGKGKGEKGTKVGADDTNGTDPDPEPEPEPEPEPEPEPEPE--- 81

Qy      96 MDPPNPEEVKTPEEG 110
              | ||    || |
Db      82 ---PEPEPEPEPG 93


RESULT   7
THYA_MOUSE
ID   THYA_MOUSE     STANDARD;      PRT;   110 AA.
AC   P26350;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Prothymosin alpha.
GN   PTMA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91198154; PubMed=2015308;
RA   Schmidt G., Werner D.;
RT   "Nucleotide sequence of the murine prothymosin alpha cDNA and its
RT   deduced primary and secondary protein structure.";
RL   Biochim. Biophys. Acta 1088:442-444(1991).
RN   [2]
RP   SEQUENCE.
RX   MEDLINE=91032153; PubMed=2226839;
RA   Low T.L.K., Pan T.L., Lin Y.S.;
RT   "Depression of prothymosin alpha production in murine thymus
RT   correlates with staphylococcal enterotoxin-B-induced
RT   immunosuppression.";
RL   FEBS Lett. 273:1-5(1990).
RN   [3]
```

```
RP   SEQUENCE OF 14-17; 21-67 AND 90-102.
RX   MEDLINE=90060328; PubMed=2479575;
RA   Makarova T., Grebenshikov N., Egorov C., Vartapetian A., Bogdanov A.;
RT   "Prothymosin alpha is an evolutionary conserved protein covalently
RT   linked to a small RNA.";
RL   FEBS Lett. 257:247-250(1989).
CC   -!- FUNCTION: Prothymosin alpha may mediate immune function by
CC       conferring resistance to certain opportunistic infections.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: COVALENTLY LINKED TO A SMALL RNA OF ABOUT 20 NUCLEOTIDES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X56135; CAA39601.1; -.
DR   PIR; S06666; S06666.
DR   PIR; S12883; S12883.
DR   PIR; S15073; S15073.
DR   MGD; MGI:97803; Ptma.
KW   Acetylation; Nuclear protein.
FT   INIT_MET      0      0
FT   PEPTIDE       1     28       THYMOSIN ALPHA.
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN       41    100       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT    105    105       K -> KK (IN REF. 2).
SQ   SEQUENCE   110 AA;  12123 MW;  AF321A324731826F CRC64;


  Query Match             9.6%;  Score 62;  DB 1;  Length 110;
  Best Local Similarity   24.7%;  Pred. No. 1.8e+02;
  Matches   19;  Conservative   13;  Mismatches   43;  Indels    2;  Gaps    1;

Qy      35 VEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQ 94
           || |  | |  |    :    : || |  :     : |   :|:   :   | ||| : :
Db      23 VEEA--ENGRDAPANGNAQNEENGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDE 80

Qy      95 EMDPPNPEEVKTPEEGE 111
           | : |  : |   :| :
Db      81 EAEAPTGKRVAEDDEDD 97


RESULT   8
THYA_RAT
ID   THYA_RAT       STANDARD;      PRT;   111 AA.
AC   P06302;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Prothymosin alpha.
GN   PTMA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=88239673; PubMed=3377505;
RA   Frangou-Lazaridis M., Clinton M., Goodall G.J., Horecker B.L.;
RT   "Prothymosin alpha and parathymosin: amino acid sequences deduced
RT   from the cloned rat spleen cDNAs.";
RL   Arch. Biochem. Biophys. 263:305-310(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91114696; PubMed=1989881;
RA   Eilers M., Schirm S., Bishop J.M.;
RT   "The MYC protein activates transcription of the alpha-prothymosin
RT   gene.";
RL   EMBO J. 10:133-141(1991).
RN   [3]
RP   PRELIMINARY SEQUENCE.
RX   MEDLINE=85113197; PubMed=3855555;
RA   Haritos A.A., Blacher R., Stein S., Caldarella J., Horecker B.L.;
RT   "Primary structure of rat thymus prothymosin alpha.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:343-346(1985).
RN   [4]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=91099511; PubMed=2269362;
RA   Palvimo J., Linnala-Kankkunen A.;
RT   "Identification of a low-Mr acidic nuclear protein as prothymosin
RT   alpha.";
RL   FEBS Lett. 277:257-260(1990).
CC   -!- FUNCTION: Prothymosin alpha may mediate immune function by
CC       conferring resistance to certain opportunistic infections.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: Covalently linked to a small RNA of about 20 nucleotides (By
CC       similarity).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20035; AAA42241.1; -.
DR   EMBL; M86564; AAA40758.1; -.
DR   EMBL; M33024; AAA41931.1; -.
DR   EMBL; X55326; CAA39028.1; -.
DR   PIR; A01519; TNRTA.
DR   PIR; A32265; A32265.
DR   PIR; S13014; S13014.
KW   Acetylation; Nuclear protein.
FT   INIT_MET      0      0
FT   PEPTIDE       1     28       THYMOSIN ALPHA.
FT   MOD_RES       1      1       ACETYLATION.
FT   DOMAIN       41    100       ASP/GLU-RICH (ACIDIC).
SQ   SEQUENCE   111 AA;  12251 MW;  39FF33AADDE5A182 CRC64;


  Query Match             9.6%;  Score 62;  DB 1;  Length 111;
  Best Local Similarity   24.7%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative   13;  Mismatches   43;  Indels    2;  Gaps    1;

Qy      35 VEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQ 94
           || |  | |  |    :    : || |  :     : |   :|:   :   | ||| : :
Db      23 VEEA--ENGRDAPANGNAQNEENGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDE 80

Qy      95 EMDPPNPEEVKTPEEGE 111
           | : |  : |   :| :
Db      81 EAEAPTGKRVAEDDEDD 97


RESULT   9
THYP_RAT
ID   THYP_RAT       STANDARD;      PRT;   101 AA.
AC   P04550;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Parathymosin (Zinc-binding 11.5 kDa protein).
GN   PTMS OR ZNBP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89338759; PubMed=2759245;
RA   Trompeter H.-I., Brand I.A., Soeling H.-D.;
```

```
RT   "The primary sequence of the PFK-1 inactivating zinc-binding protein
RT   as deduced from cDNA sequencing. Identity of the zinc-binding protein
RT   with rat parathymosin.";
RL   FEBS Lett. 253:63-66(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISCHER; TISSUE=Liver;
RX   MEDLINE=92183876; PubMed=1544455;
RA   Trompeter H.-I., Soeling H.-D.;
RT   "Cloning and characterisation of a gene encoding the 11.5 kDa zinc-
RT   binding protein (parathymosin-alpha).";
RL   FEBS Lett. 298:245-248(1992).
RN   [3]
RP   SEQUENCE OF 22-101 FROM N.A.
RC   TISSUE=Spleen;
RX   MEDLINE=88239673; PubMed=3377505;
RA   Frangou-Lazaridis M., Clinton M., Goodall G.J., Horecker B.L.;
RT   "Prothymosin alpha and parathymosin: amino acid sequences deduced
RT   from the cloned rat spleen cDNAs.";
RL   Arch. Biochem. Biophys. 263:305-310(1988).
RN   [4]
RP   PRELIMINARY SEQUENCE.
RX   MEDLINE=86149281; PubMed=3456585;
RA   Komiyama T., Pan L.-X., Haritos A.A., Wideman J.W., Pan Y.-C.E.,
RA   Chang M., Rogers I., Horecker B.L.;
RT   "The primary structure of rat parathymosin.";
RL   Proc. Natl. Acad. Sci. U.S.A. 83:1242-1245(1986).
RN   [5]
RP   PRELIMINARY SEQUENCE OF 1-30.
RX   MEDLINE=85140226; PubMed=3856246;
RA   Haritos A.A., Salvin S.B., Blacher R., Stein S., Horecker B.L.;
RT   "Parathymosin alpha: a peptide from rat tissues with structural
RT   homology to prothymosin alpha.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:1050-1053(1985).
RN   [6]
RP   REVISION TO 1-4.
RX   MEDLINE=88339930; PubMed=3421960;
RA   Panneerselvam C., Clinton M., Wellner D., Horecker B.L.;
RT   "Bovine parathymosin: amino acid sequence and comparison with rat
RT   parathymosin.";
RL   Biochem. Biophys. Res. Commun. 155:539-545(1988).
CC   -!- FUNCTION: Parathymosin may mediate immune function by blocking
CC       the effect of prothymosin alpha which confers resistance to
CC       certain opportunistic infections.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M20616; AAA42249.1; -.
DR   EMBL; M33025; AAA41810.1; -.
DR   EMBL; X64053; CAA45411.1; ALT_SEQ.
DR   EMBL; X16481; CAA34501.1; -.
DR   PIR; B31512; B31512.
DR   PIR; A05095; A05095.
DR   PIR; A22075; A22075.
DR   PIR; B32265; B32265.
DR   PIR; S05212; S05212.
KW   Immune response; Acetylation; Zinc.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION.
FT   DOMAIN       35     89       ASP/GLU-RICH (ACIDIC).
SQ   SEQUENCE   101 AA;  11428 MW;  D32BE03032BD85D4 CRC64;


  Query Match             9.5%;  Score 61.5;  DB 1;  Length 101;
  Best Local Similarity   25.7%;  Pred. No. 1.9e+02;
  Matches   19;  Conservative   15;  Mismatches   23;  Indels   17;  Gaps    4;

Qy      41 EEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGP----DGQEM 96
           || |   : ::   |::|||:     :| : :   :|        | ::||      :|
Db      38 EEEENGAEEEEEETAEDGEDD----DEGDEEDEEEEE--------EEDEGPVRKRTAEEE 85

Qy      97 DPPNPEEVKTPEEG 110
           |  :|:  || | |
Db      86 DEADPKRQKT-ENG 98


RESULT  10
MEA1_PIG
ID   MEA1_PIG       STANDARD;      PRT;   113 AA.
AC   Q95313;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Male-enhanced antigen-1 (Mea-1) (Fragment).
GN   MEA1 OR MEA.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Small intestine;
RA   Winteroe A.K., Fredholm M., Davies W.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
CC       TESTIS DEVELOPMENT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z81165; CAB03550.1; -.
KW   Spermatogenesis; Developmental protein.
FT   NON_TER     113    113
SQ   SEQUENCE   113 AA;  12083 MW;  1F52DC82FDAA95B6 CRC64;


  Query Match             9.4%;  Score 61;  DB 1;  Length 113;
  Best Local Similarity   25.0%;  Pred. No. 2.2e+02;
  Matches   27;  Conservative   17;  Mismatches   40;  Indels   24;  Gaps    6;

Qy      21 EMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGH 80
           :::|| |   | :: |   | :|   |:     :::| |:|    | ||   ::
Db       9 DIMGPER--IFPNQTEELGPHQG--PTEGTGDWSSEEPEEEQEETGTGPAGYSYQPLNQD 64

Qy      81 PQTGCECEDGP--DGQEM---------------DPPNPEEVKTPEEG 110
           |:   | |  |  ||:::                |||   |    |||
Db      65 PEQE-EVELAPVGDGEDVVADIQDRIQALGLHLPDPPLESE-DEDEEG 110


RESULT  11
PARA_TRYBB
ID   PARA_TRYBB     STANDARD;      PRT;   114 AA.
AC   P18764;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Procyclic form specific polypeptide A-alpha precursor (Procyclin A-
DE   alpha) (PARP A-alpha).
GN   PARPA-ALPHA.
OS   Trypanosoma brucei brucei.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX   NCBI_TaxID=5702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89261740; PubMed=2725502;
```

```
RA   Mowatt M.R., Wisdom G.S., Clayton C.E.;
RT   "Variation of tandem repeats in the developmentally regulated
RT   procyclic acidic repetitive proteins of Trypanosoma brucei.";
RL   Mol. Cell. Biol. 9:1332-1335(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90258895; PubMed=2342468;
RA   Clayton C.E., Fueri J.P., Itzhaki J.E., Bellofatto V., Sherman D.R.,
RA   Wisdom G.S., Vijayasarathy S., Mowatt M.R.;
RT   "Transcription of the procyclic acidic repetitive protein genes of
RT   Trypanosoma brucei.";
RL   Mol. Cell. Biol. 10:3036-3047(1990).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=427;
RA   Vijayasarathy S., Ernest I., Itzhaki J., Sherman D., Mowatt M.R.,
RA   Michels P.A.M., Clayton C.E.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAJOR SURFACE ANTIGEN OF PROCYCLIC FORMS.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED ONLY AT A CERTAIN STAGE DURING
CC       DIFFERENTIATION IN THE INSECT VECTOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M25787; AAA53283.1; -.
DR   EMBL; M33129; AAA30224.1; -.
DR   EMBL; X52584; CAA36814.1; -.
DR   PIR; A30229; A30229.
DR   PIR; S21539; S21539.
KW   Signal; Antigen; Repeat; GPI-anchor.
FT   SIGNAL        1     27
FT   CHAIN        28     92       PROCYCLIC FORM SPECIFIC POLYPEPTIDE A-
FT                                ALPHA.
FT   PROPEP       93    114
FT   DOMAIN       48     85       6 X 5 AA TANDEM REPEATS OF G-P-E-E-T.
FT   LIPID        92     92       GPI-ANCHOR.
SQ   SEQUENCE   114 AA;  11611 MW;  FFF2690DAAAE445E CRC64;
```

```
  Query Match             9.4%;  Score 61;  DB 1;  Length 114;
  Best Local Similarity   34.4%;  Pred. No. 2.2e+02;
  Matches   22;  Conservative    8;  Mismatches   24;  Indels   10;  Gaps    4;

Qy      60 DEGAS-----AGQGPKPE---AHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGE 111
           || ||      |:| : |      :| |  :|| | | ||:    :   |||   ||| |
Db      29 DESASNVIVKGGKGKEREDGPEEPEETGPEETGPE-ETGPEETGPEETGPEET-GPEETE 86

Qy     112 KQSQ 115
           : :
Db      87 PEPE 90
```

```
RESULT  12
PGCA_PIG
ID   PGCA_PIG       STANDARD;      PRT;    83 AA.
AC   Q29011;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Aggrecan core protein (Cartilage-specific proteoglycan core protein)
DE   (CSPCP) (Fragment).
GN   AGC1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=95128522; PubMed=7827755;
RA   Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT   "Length variation in the keratan sulfate domain of mammalian
RT   aggrecan.";
RL   Matrix Biol. 14:323-328(1994).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC       MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC       IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC       HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC       REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX (BY
CC       SIMILARITY).
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S74664; AAC60528.1; -.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR001304; lectin_c.
DR   PROSITE; PS01241; LINK; PARTIAL.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; PARTIAL.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; PARTIAL.
KW   Glycoprotein; Cartilage; Proteoglycan; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1     >83       KS.
FT   NON_TER      83      83
SQ   SEQUENCE   83 AA;  8576 MW;  99DC5641CBCD44A8 CRC64;
```

```
  Query Match             9.4%;  Score 60.5;  DB 1;  Length 83;
  Best Local Similarity   27.3%;  Pred. No. 1.8e+02;
  Matches   24;  Conservative    9;  Mismatches   34;  Indels   21;  Gaps    4;

Qy      24 GPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQT 83
           ||   | |   ||   |: |:   ::|  ::|   |  || : | |     ||  | |
Db      10 GPSATEAPSTSEEPFPSEKPFPS---EEPFPSEEFPSEKPSASEEPFP----SEQ--PST 60

Qy      84 GCECEDGPDGQEMDPPNPEEVKTPEEGE 111
              :  | |   : |  ||
Db      61 -----------LSAPVPSRTELPGSGE 76
```

```
RESULT  13
THYP_HUMAN
ID   THYP_HUMAN     STANDARD;      PRT;   101 AA.
AC   P20962;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Parathymosin.
GN   PTMS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=89149806; PubMed=2537638;
RA   Clinton M., Frangou-Lazaridis M., Panneerselvam C., Horecker B.L.;
RT   "The sequence of human parathymosin deduced from a cloned human
RT   kidney cDNA.";
```

```
RL   Biochem. Biophys. Res. Commun. 158:855-862(1989).
CC   -!- FUNCTION: Parathymosin may mediate immune function by blocking
CC       the effect of prothymosin alpha which confers resistance to
CC       certain opportunistic infections.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M24398; AAA61185.1; -.
DR   PIR; A32264; A32264.
DR   MIM; 168440; -.
KW   Immune response; Acetylation.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       ACETYLATION (BY SIMILARITY).
FT   DOMAIN       35     89       ASP/GLU-RICH (ACIDIC).
SQ   SEQUENCE   101 AA;  11399 MW;  D42A6C04B25741C6 CRC64;


  Query Match             9.4%;  Score 60.5;  DB 1;  Length 101;
  Best Local Similarity   25.7%;  Pred. No. 2.2e+02;
  Matches   19;  Conservative   15;  Mismatches   23;  Indels   17;  Gaps    4;

Qy      41 EEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGP----DGQEM 96
           || |   : ::   |::||:|     :| : :   :|        | ::||      :|
Db      38 EEEENGAEEEEETAEDGEEE----DEGEEEDEEEEE--------EDDEGPALKRAAEEE 85

Qy      97 DPPNPEEVKTPEEG 110
           |  :|:  || | |
Db      86 DEADPKRQKT-ENG 98


RESULT  14
THYA_BOVIN
ID   THYA_BOVIN     STANDARD;      PRT;   109 AA.
AC   P01252;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Prothymosin alpha.
GN   PTMA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=88339364; PubMed=2901823;
RA   Panneerselvam C., Wellner D., Horecker B.L.;
RT   "The amino acid sequence of bovine thymus prothymosin alpha.";
RL   Arch. Biochem. Biophys. 265:454-457(1988).
RN   [2]
RP   SEQUENCE OF 1-28.
RX   MEDLINE=79109664; PubMed=762108;
RA   Low T.L.K., Goldstein A.L.;
RT   "The chemistry and biology of thymosin. II. Amino acid sequence
RT   analysis of thymosin alpha1 and polypeptide beta1.";
RL   J. Biol. Chem. 254:987-995(1979).
RN   [3]
RP   PHOSPHORYLATION BY CK2.
RX   MEDLINE=93050183; PubMed=1426245;
RA   Barcia M.G., Castro J.M., Jullien C.D., Gonzalez C.G., Freire M.;
RT   "Prothymosin alpha is phosphorylated by casein kinase-2.";
RL   FEBS Lett. 312:152-156(1992).
CC   -!- FUNCTION: Prothymosin alpha may mediate immune function by
CC       conferring resistance to certain opportunistic infections.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- PTM: Covalently linked to a small RNA of about 20 nucleotides (By
CC       similarity).
DR   PIR; A01520; TNBOA1.
DR   PIR; S06426; S06426.
KW   Acetylation; Nuclear protein; Phosphorylation.
FT   PEPTIDE       1     28       THYMOSIN ALPHA-1.
FT   MOD_RES       1      1       ACETYLATION.
FT   DOMAIN       40     99       ASP/GLU-RICH (ACIDIC).
FT   MOD_RES       7      7       PHOSPHORYLATION (BY CK2).
FT   MOD_RES      12     12       PHOSPHORYLATION (BY CK2).
FT   MOD_RES      13     13       PHOSPHORYLATION (BY CK2).
SQ   SEQUENCE   109 AA;  11941 MW;  ECB151F0A9F99D24 CRC64;


  Query Match             9.2%;  Score 59.5;  DB 1;  Length 109;
  Best Local Similarity   23.0%;  Pred. No. 2.7e+02;
  Matches   17;  Conservative   13;  Mismatches   39;  Indels    5;  Gaps    2;

Qy      38 ATPEEGEPATQRQDPAAAQEG--EDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQE 95
           |  | ||      :     :|| |:|    | | : :    |:    ||    : : :
Db      38 ANEENGEQEADNEVDEEEEEGGEEEEEEEEGDGEEEDGDEDEEAEAATGKRAAEDDEDDD 97

Qy      96 MDPPNPEEVKTPEE 109
           :|    :: || |:
Db      98 VD---TKKQKTDED 108


RESULT  15
RL12_HALHA
ID   RL12_HALHA     STANDARD;      PRT;   114 AA.
AC   P05768;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   50S ribosomal protein L12P ('A' type) (HL20).
GN   RPL12P.
OS   Halobacterium halobium, and
OS   Halobacterium cutirubrum.
OC   Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC   Halobacterium.
OX   NCBI_TaxID=2242, 2242;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=H.halobium; STRAIN=A9;
RA   Itoh T., Kumazaki T., Sugiyama M., Otaka E.;
RT   "Molecular cloning and sequence analysis of the ribosomal 'A' protein
RT   gene from the archaebacterium, Halobacterium halobium.";
RL   Biochim. Biophys. Acta 949:110-118(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=H.halobium; STRAIN=S9;
RX   MEDLINE=88329082; PubMed=2458258;
RA   Itoh T.;
RT   "Complete nucleotide sequence of the ribosomal 'A' protein operon
RT   from the archaebacterium, Halobacterium halobium.";
RL   Eur. J. Biochem. 176:297-303(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=H.cutirubrum; STRAIN=34001;
RX   MEDLINE=89305527; PubMed=2743981;
RA   Shimmin L.C., Dennis P.P.;
RT   "Characterization of the L11, L1, L10 and L12 equivalent ribosomal
RT   protein gene cluster of the halophilic archaebacterium Halobacterium
RT   cutirubrum.";
RL   EMBO J. 8:1225-1235(1989).
RN   [4]
RP   SEQUENCE.
RC   SPECIES=H.cutirubrum;
RA   Oda G.;
RL   Unpublished results, cited by:
RL   Wittmann-Liebold B.;
RL   (In) Hardesty B., Kramer G. (eds.);
RL   Structure, function and genetics of ribosomes, pp.326-361,
RL   Springer-Verlag, New York (1986).
```

```
RN   [5]
RP   SEQUENCE OF 1-76.
RC   SPECIES=H.cutirubrum;
RA   Yaguchi M., Matheson A.T., Visentin L.P., Zuker M.;
RL   (In) Osawa S., Ozeki H., Uchida H., Yura T. (eds.);
RL   Genetics and evolution of RNA polymerase, tRNA and ribosomes,
RL   pp.585-599, University of Tokyo Press, Tokyo (1980).
CC   -!- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
CC       INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
CC       ACCURATE TRANSLATION.
CC   -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X06736; CAA29915.1; -.
DR   EMBL; X13008; CAA31432.1; -.
DR   EMBL; X15078; CAA33181.1; -.
DR   PIR; S01745; R5HS2H.
DR   PIR; S04121; R5HS2C.
DR   InterPro; IPR001813; 60s_ribosomal.
DR   Pfam; PF00428; 60s_ribosomal; 1.
KW   Ribosomal protein.
SQ   SEQUENCE   114 AA;  11562 MW;  CBEB2E38BD4A644C CRC64;
```

Query Match 9.2%; Score 59.5; DB 1; Length 114;
Best Local Similarity 24.6%; Pred. No. 2.8e+02;
Matches 16; Conservative 7; Mismatches 27; Indels 15; Gaps 1;

```
Qy      34 EVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDG 93
           |   | |     |:   | ||| :|:|:               |:       | ||  |
Db      55 EEAAAAPAAAPAASGSDDEAAADDGDDD--------------EEADADEAAEAEDAGDD 99

Qy      94 QEMDP 98
             : :|
Db     100 DDEEP 104
```

Search completed: July 1, 2002, 06:39:48
Job time: 189 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:37:09 ; Search time 21.81 Seconds
(without alignments)
920.101 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 133496

Minimum DB seq length: 0
Maximum DB seq length: 116

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	637	98.6	116	4	Q9UEU5	Q9ueu5 homo sapien
2	229.5	35.5	111	4	Q96GT9	Q96gt9 homo sapien
3	107.5	16.6	110	4	Q96GU1	Q96gu1 homo sapien
4	84.5	13.1	90	5	O18562	O18562 strongyloid
5	82.5	12.8	108	5	Q9GYX8	Q9gyx8 fasciola he
6	79	12.2	90	12	P87914	P87914 human herpe
7	76	11.8	74	11	Q9QZK1	Q9qzk1 cavia porce
8	75	11.6	81	12	Q85063	Q85063 cottontail
9	74.5	11.5	91	12	Q66642	Q66642 equine herp
10	72	11.1	93	4	Q15215	Q15215 homo sapien
11	71.5	11.1	99	5	Q23793	Q23793 chironomus
12	71.5	11.1	107	5	Q9N3B4	Q9n3b4 caenorhabdi
13	71	11.0	75	2	Q9KIL8	Q9kil8 streptomyce
14	71	11.0	114	2	Q9KYF6	Q9kyf6 streptomyce
15	70.5	10.9	100	4	Q9Y3S3	Q9y3s3 homo sapien
16	70.5	10.9	114	4	Q9H1E6	Q9h1e6 homo sapien
17	70	10.8	101	17	Q9YEP0	Q9yep0 aeropyrum p
18	69.5	10.8	91	6	O18928	O18928 macaca radi
19	69.5	10.8	99	5	Q23794	Q23794 chironomus
20	69	10.7	115	4	Q9ULQ8	Q9ulq8 homo sapien
21	68	10.5	75	2	Q9RKB8	Q9rkb8 streptomyce
22	68	10.5	82	5	Q9VUQ5	Q9vuq5 drosophila
23	67.5	10.4	98	10	O80818	O80818 arabidopsis
24	67	10.4	95	6	Q9GKX4	Q9gkx4 canis famil
25	66.5	10.3	107	4	Q15254	Q15254 homo sapien
26	66	10.2	92	12	Q87076	Q87076 pseudorabie
27	66	10.2	99	5	Q9VPS9	Q9vps9 drosophila
28	65.5	10.1	66	4	Q13902	Q13902 homo sapien
29	64	9.9	82	6	Q9TV59	Q9tv59 canis famil
30	64	9.9	106	4	Q9GZP8	Q9gzp8 homo sapien
31	63.5	9.8	82	6	Q9TV58	Q9tv58 canis famil
32	63.5	9.8	105	4	Q9UED2	Q9ued2 homo sapien
33	63.5	9.8	109	11	Q923P0	Q923p0 mus musculu
34	63.5	9.8	116	6	Q9TV57	Q9tv57 canis famil
35	63	9.8	105	4	Q9NYD3	Q9nyd3 homo sapien
36	63	9.8	107	5	Q9VPS8	Q9vps8 drosophila
37	62.5	9.7	99	6	Q9GKX5	Q9gkx5 canis famil
38	62.5	9.7	101	4	Q9UMZ1	Q9umz1 homo sapien
39	62	9.6	101	11	Q9D0J8	Q9d0j8 mus musculu
40	62	9.6	108	4	Q15202	Q15202 homo sapien
41	62	9.6	110	10	Q9FH00	Q9fh00 arabidopsis
42	62	9.6	115	12	Q84278	Q84278 canine oral
43	62	9.6	116	17	Q9YB29	Q9yb29 aeropyrum p
44	61.5	9.5	53	13	Q9DFL0	Q9df10 gillichthys
45	61.5	9.5	106	13	Q98UF2	Q98uf2 xenopus lae

ALIGNMENTS

RESULT 1
Q9UEU5
ID Q9UEU5 PRELIMINARY; PRT; 116 AA.
AC Q9UEU5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GAGE-8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czekay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
RT human cancers and in normal testis.";
RL Cancer Res. 59:3157-3165(1999).
DR EMBL; AF055473; AAC33676.1; -.
SQ SEQUENCE 116 AA; 12764 MW; DD3052939E66F19A CRC64;

Query Match 98.6%; Score 637; DB 4; Length 116;
Best Local Similarity 99.1%; Pred. No. 5.1e-50;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      61 EGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
```

RESULT 2

```
Q96GT9
ID   Q96GT9      PRELIMINARY;      PRT;   111 AA.
AC   Q96GT9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO G ANTIGEN 8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA, AND CHORIOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009232; AAH09232.1; -.
SQ   SEQUENCE   111 AA;  12354 MW;  38C93332C5BA0E14 CRC64;


  Query Match             35.5%;  Score 229.5;  DB 4;  Length 111;
  Best Local Similarity   45.8%;  Pred. No. 1.5e-13;
  Matches   54;  Conservative   15;  Mismatches   38;  Indels   11;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQEG 58
           |||||||||||||||| ::|||:|| |        :||    |:| :| |: ::|   |:
Db       1 MSWRGRSTYRPRPRRSLQPPELIGAM--------LEPTDEEPKEEKPPTKSRNPTPDQKR 52

Qy      59 E-DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
           | |:||:  | |  ||  ||    :||  || | | :    |  |  | || || :||
Db      53 EDDQGAAEIQVPDLEADLQELCQTKTGDGCEGGTDVKGKILPKAEHFKMPEAGEGKSQ 110


RESULT   3
Q96GU1
ID   Q96GU1      PRELIMINARY;      PRT;   110 AA.
AC   Q96GU1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:16481).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009230; AAH09230.1; -.
SQ   SEQUENCE   110 AA;  11777 MW;  6D6AF4563135BB6C CRC64;


  Query Match             16.6%;  Score 107.5;  DB 4;  Length 110;
  Best Local Similarity   32.7%;  Pred. No. 0.012;
  Matches   36;  Conservative   16;  Mismatches   51;  Indels    7;  Gaps    2;

Qy       4 RGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGA 63
           | :|: |   :   :|   :||:  :| ::|      :| || |  |  | : | ::|||
Db       7 RSQSSERGNDQESSQP---VGPVIVQQPTEEKR----QEEEPPTDNQGIAPSGEIKNEGA 59

Qy      64 SAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           | ||   ||  ||    :      |||| :|   |  :  |  | || |
Db      60 PAVQGTDVEAFQQELALLKIEDAPGDGPDVREGTLPTFDPTKVLEAGEGQ 109


RESULT   4
O18562
ID   O18562      PRELIMINARY;      PRT;    90 AA.
AC   O18562;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE   WITH STRONGYLOIDIASIS (FRAGMENT).
OS   Strongyloides stercoralis.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Strongyloididae; Strongyloides.
OX   NCBI_TaxID=6248;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ramachandran S., Thompson R.W., Gam A.A., Neva F.A.;
RT   "A set of recombinant clones for immunodiagnosis of
RT   strongyloidiasis.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U90352; AAB65140.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   90 AA;  9438 MW;  D2373206920E68B4 CRC64;


  Query Match             13.1%;  Score 84.5;  DB 5;  Length 90;
  Best Local Similarity   24.8%;  Pred. No. 1.1;
  Matches   25;  Conservative   16;  Mismatches   33;  Indels   27;  Gaps    4;

Qy      10 RPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGP 69
           :|:|:   : ||              ||  ||| |   : : |   :  | || :  : |
Db       9 KPKPKTTTQAPE-------------EPEGPEEPEGPEEPEGPEGPE--EPEGPAGPEEP 52

Qy      70 KPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEG 110
           :  |  :|   |:        |:|    |  ||| :  ::|
Db      53 EGPAGPEEPEGPEE-------PEG----PAGPEEPRDDDDG 82


RESULT   5
Q9GYX8
ID   Q9GYX8      PRELIMINARY;      PRT;   108 AA.
AC   Q9GYX8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TEGUMENTAL ANTIGEN-LIKE PROTEIN (FRAGMENT).
OS   Fasciola hepatica (Liver fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
OC   Echinostomida; Echinostomata; Fascioloidea; Fasciolidae; Fasciola.
OX   NCBI_TaxID=6192;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Toledo R., Rodriguez-Medina J.R., Hillyer G.V.;
RT   "Molecular cloning and characterization of a Fasciola hepatica gene
RT   encoding a proline-rich polypeptide related to tegumental antigens.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF286861; AAG00994.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
FT   NON_TER       1       1
SQ   SEQUENCE   108 AA;  11325 MW;  9C208DDE3F050055 CRC64;


  Query Match             12.8%;  Score 82.5;  DB 5;  Length 108;
  Best Local Similarity   32.2%;  Pred. No. 2;
  Matches   28;  Conservative    6;  Mismatches   34;  Indels   19;  Gaps    4;

Qy      24 GPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQT 83
           || :||     ||     ||  : :||    | |         |||          |:
Db      12 GP-KPEDPEPGPEPEPKPGPEPGPKPEDPEPGPEPE---------PKP--------GPEP 53

Qy      84 GCECEDGPDGQEMDP-PNPEEVKTPEE 109
           | : ||    | | :| | ||    |||
Db      54 GPKPEDPEPGPEPEPKPGPEPGPKPEE 80


RESULT   6
P87914
ID   P87914      PRELIMINARY;      PRT;    90 AA.
AC   P87914;
```

```
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ICP0D.
GN   ALPHA 0.
OS   human herpesvirus 1.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=10298;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F;
RA   Carter K.L., Roizman B.;
RT   "Alternatively spliced mRNAs predicted to yield frame-shift genes and
RT   stable intron 1 RNAs of the herpes simplex virus 1 regulatory gene
RT   alpha 0 accumulate in the cytoplasm of infected cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1996).
DR   EMBL; U70040; AAB37297.1; -.
SQ   SEQUENCE   90 AA;  9210 MW;  9B0562ECF6D55C8C CRC64;


  Query Match             12.2%;  Score 79;  DB 12;  Length 90;
  Best Local Similarity   29.0%;  Pred. No. 3.5;
  Matches   29;  Conservative    8;  Mismatches   25;  Indels   38;  Gaps    5;

Qy       5 GRSTYRP--RPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEG 62
           | || ||  ||:|  |||  : |  |               | |             . | |
Db       6 GASTRRPEGRPQR-EEPPAIFGGRR--------------GTP-------------DSG 35

Qy      63 ASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPE 102
           || |:  : |  |  :|  ||   |      : : ||: |
Db      36 ASGGREKREEEGSGSKGRTQTTFGC------RPLSPPSSE 69


RESULT   7
Q9QZK1
ID   Q9QZK1      PRELIMINARY;      PRT;    74 AA.
AC   Q9QZK1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SPERAD-7.
OS   Cavia porcellus (Guinea pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX   NCBI_TaxID=10141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ilayperuma I., Allen C.A., Gill D., Canas B., Hessian P.A., Pappin D.,
RA   Stanton J.L., Fisher L.C., Banhalmi R., Green D.P.L.;
RT   "Sperad is present in the equatorial segment of guinea pig sperm and
RT   may play a role in sperm-egg adhesion and egg activation.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF180148; AAD56014.1; -.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
SQ   SEQUENCE   74 AA;  8060 MW;  6B19E54D5ED90D27 CRC64;


  Query Match             11.8%;  Score 76;  DB 11;  Length 74;
  Best Local Similarity   32.9%;  Pred. No. 5.3;
  Matches   26;  Conservative    7;  Mismatches   34;  Indels   12;  Gaps    4;

Qy       6 RSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASA 65
           | |   :|:  || |  | |  |   :   ||  ||  :|    | |   |  : |
Db       4 RKTGGEQPK--VEAPAEIKPTEP---TGPTEPPQPE--QPPQPEQPPQPEQPPQPE---- 52

Qy      66 GQGPKPEAHSQEQGHPQTG 84
             | |:||   | :  || |
Db      53 -QPPQPEQPPQPEQPPQPG 70


RESULT   8
```

```
Q85063
ID   Q85063      PRELIMINARY;      PRT;    81 AA.
AC   Q85063;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PAPILLOMAVIRUS SYLVILAGI WASHINGTON B (E2) AND (E4) GENES
DE   (FRAGMENT).
GN   E4.
OS   Cottontail rabbit papillomavirus.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WASHINGTON B;
RA   Wu X., Brandsma J.;
RT   "Papilloma Formation by Cottontail Rabbit Papillomavirus (CRPV)
RT   Requires E1 and E2 Regulatory Genes in Addition to the E6 and E7
RT   Transforming Genes.";
RL   J. Virol. 0:0-0(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WASHINGTON B;
RX   MEDLINE=85166175; PubMed=2984661;
RA   Giri I., Danos O., Yaniv M.;
RT   "Genomic structure of the cottontail rabbit (Shope) papillomavirus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:1580-1584(1985).
DR   EMBL; U09497; AAB60544.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   81 AA;  9173 MW;  4E80200A0004DF5C CRC64;


  Query Match             11.6%;  Score 75;  DB 12;  Length 81;
  Best Local Similarity   30.6%;  Pred. No. 7.1;
  Matches   15;  Conservative    7;  Mismatches   25;  Indels    2;  Gaps    1;

Qy      67 QGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
           |||||  |  ::|    |  | :|    |  ||  : :  |   ::  |
Db      11 QGPKPRVHWADEGQGHQG--CNEGRQSNENRPPRTKRILLPGTSDRLLQ 57


RESULT   9
Q66642
ID   Q66642      PRELIMINARY;      PRT;    91 AA.
AC   Q66642;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   MYRISTYLATED TEGUMENT PROTEIN.
OS   Equine herpesvirus type 2 (strain 86/87) (EHV-2).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae.
OX   NCBI_TaxID=82831;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95302501; PubMed=7783207;
RA   Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;
RT   "The DNA sequence of equine herpesvirus 2.";
RL   J. Mol. Biol. 249:520-528(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Telford E.A.R.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U20824; AAC13826.1; -.
SQ   SEQUENCE   91 AA;  10311 MW;  1784D979FAA61F40 CRC64;


  Query Match             11.5%;  Score 74.5;  DB 12;  Length 91;
  Best Local Similarity   30.8%;  Pred. No. 8.9;
  Matches   24;  Conservative    9;  Mismatches   38;  Indels    7;  Gaps    2;

Qy      36 EPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQE 95
```

```
          :| |  ||:|   |::    :|| ||   |     |   : |       | ::| || |
Db      13 DPLTDVEGQPINVREEFEMFEEG-DEATEACVFLNPNMGTDEYD------EEDEGGDGGE 65

Qy      96 MDPPNPEEVKTPEEGEKQ 113
           | ||   ||    |:
Db      66 GREPQPEVKTTPYPKRKK 83


RESULT  10
Q15215
ID   Q15215      PRELIMINARY;      PRT;    93 AA.
AC   Q15215;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SALIVARY PROLINE-RICH PROTEIN 1 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84298176; PubMed=6089212;
RA   Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,
RA   Maeda N., Vanin E.F., Carlson D.M., Smithies O.;
RT   "Clones from the human gene complex coding for salivary proline-rich
RT   proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).
DR   EMBL; K02576; AAA36503.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   93 AA;  9170 MW;  04AD5F0797E31867 CRC64;


  Query Match             11.1%;  Score 72;  DB 4;  Length 93;
  Best Local Similarity   25.3%;  Pred. No. 15;
  Matches   19;  Conservative    9;  Mismatches   19;  Indels   28;  Gaps    3;

Qy      67 QGPKPEAHSQEQGHP----------QTGCECED--GPDGQEMDPP--------------- 99
           ||| |   :: || |          | ||: :    | |:   ||
Db       3 QGPPPTGGNKPQGPPPPGKPQGPPAQGGCKSQSARAPPGKPQGPPQQEGNNPQGPPPPAG 62

Qy     100 -NPEEVKTPEEGEKQ 113
            ||:: : |  |: |
Db      63 GNPQQPQAPPAGQPQ 77


RESULT  11
Q23793
ID   Q23793      PRELIMINARY;      PRT;    99 AA.
AC   Q23793;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HIGH MOBILITY GROUP PROTEIN I/Y.
OS   Chironomus tentans (Midge).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
OC   Chironomoidea; Chironomidae; Chironominae; Chironomus.
OX   NCBI_TaxID=7153;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95105193; PubMed=7806532;
RA   Claus P., Schulze E., Wisniewski J.R.;
RT   "Insect proteins homologous to mammalian high mobility group proteins
RT   I/Y (HMG I/Y): Characterization and binding to linear and four-way
RT   junction DNA.";
RL   J. Biol. Chem. 269:33042-33048(1994).
DR   EMBL; Z36897; CAA85364.1; -.
DR   InterPro; IPR000637; AT_hook.
DR   Pfam; PF02178; AT_hook; 3.
DR   PRINTS; PR00929; ATHOOK.
DR   SMART; SM00384; AT_hook; 3.
SQ   SEQUENCE   99 AA;  10472 MW;  5BA44A226B8A6990 CRC64;


  Query Match             11.1%;  Score 71.5;  DB 5;  Length 99;
  Best Local Similarity   24.7%;  Pred. No. 18;
  Matches   22;  Conservative   21;  Mismatches   43;  Indels    3;  Gaps    3;

Qy      10 RPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQ-G 68
           | || :  :| |   |   ::   :||   |:: | :|: :: || ::|    :   :
Db      10 RGRPAK-AKPEETASPKAAKKEEKKVE-EVPKKIEESTKPENGAAPKKGRGRPSKGDKAA 67

Qy      69 PKPEAHSQEQGHPQTGCECEDGPDGQEMD 97
           ||  |  : :| |    :  |  | :|::
Db      68 PKRPASGKGRGRPAKNAKKVDDADAEEVN 96


RESULT  12
Q9N3B4
ID   Q9N3B4      PRELIMINARY;      PRT;   107 AA.
AC   Q9N3B4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 11.9 KDA PROTEIN.
GN   Y54G2A.19.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wollam C., Zidanic M.;
RT   "The sequence of C. elegans cosmid Y54G2A.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC024817; AAF59579.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   107 AA;  11887 MW;  5A8456C4FB5B1D85 CRC64;


  Query Match             11.1%;  Score 71.5;  DB 5;  Length 107;
  Best Local Similarity   25.8%;  Pred. No. 20;
  Matches   25;  Conservative   13;  Mismatches   32;  Indels   27;  Gaps    5;

Qy      32 SDEVEPATPEEGEPATQR------QDPAAAQEGEDEGASAGQGPKPEAH-SQEQGHPQTG 84
           : : : : |  | |:| |       ||   |  |          |: ::| |  :||| ||
Db       6 NSDSQKSLPSPGHPSTGRPDLDLITDPLEEQRLE---------PQRDSHRSDTRGHPSTG 56

Qy      85 CECEDGPDGQ------EMDPPN-----PEEVKTPEEG 110
              :  :    |        || |      ||::   ||
Db      57 RKSHNSTSAQSSVYNSRTDPENAAMDSDEELRKMMEG 93


RESULT  13
Q9KIL8
ID   Q9KIL8      PRELIMINARY;      PRT;    75 AA.
AC   Q9KIL8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
```

```
DE   F58A (HYPOTHETICAL 8.3 KDA PROTEIN).
GN   F58A OR SCP8.34C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2), AND M145;
RA   Kormanec J., Sevcikova B., Homerova D.;
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D., Harris D.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AF230490; AAF82062.1; -.
DR   EMBL; AL390975; CAC01372.1; -.
SQ   SEQUENCE   75 AA;  8263 MW;  392569EF4D29B941 CRC64;


  Query Match             11.0%;  Score 71;  DB 2;  Length 75;
  Best Local Similarity   27.8%;  Pred. No. 15;
  Matches   25;  Conservative   10;  Mismatches   35;  Indels   20;  Gaps    4;

Qy      23 IGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQ 82
           : |  || | ||  :   ||: :   :  |  ||::  | |             ||:  |
Db       3 VDPTDPETFDDEAQAPDPEDRD--VEAPDADAAEQRADVG------------QERDAPL 47

Qy      83 TGCECEDGPD-GQEMDPPNPEEVKTPEEGE 111
           || |    ||   | |     |   :| :
Db      48 TGVE----PDRADEADLAEQARVVPHDEDD 73


RESULT  14
Q9KYF6
ID   Q9KYF6      PRELIMINARY;      PRT;   114 AA.
AC   Q9KYF6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PUTATIVE SECRETED PROTEIN (FRAGMENT).
GN   SCC61A.37C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL356595; CAB92282.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   114 AA;  11673 MW;  EC396FC15B4B9991 CRC64;


  Query Match             11.0%;  Score 71;  DB 2;  Length 114;
  Best Local Similarity   29.8%;  Pred. No. 23;
  Matches   25;  Conservative    8;  Mismatches   39;  Indels   12;  Gaps    3;

Qy      32 SDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGP 91
           :|  ||    | ||       |    ||:|       | |  |  |:  | :| |    |
Db      14 NDAEEPPGYGEEEPPPPTTPPG---YGEEE-------PPPPTH--EEPPPPSGEEEPPPP 61

Qy      92 DGQEMDPPNPEEVKTPEEGEKQSQ 115
             :|  || |:  : |    |  |:
Db      62 SEEEHTPPAPQTEQPPALAETGSE 85


RESULT  15
Q9Y3S3
ID   Q9Y3S3      PRELIMINARY;      PRT;   100 AA.
AC   Q9Y3S3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   MUC-B1 PROTEIN (FRAGMENT).
GN   MUC-B1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COLON;
RA   Walter A.O., Schwaeble W.J., Thres T., Radecke K., Real F.X.,
RA   Dippold W.;
RT   "Cloning and characterisation of a novel human mucin gene with 53-
RT   base-pair tandem repeats.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ006205; CAB43492.1; -.
FT   NON_TER       1      1
FT   NON_TER     100    100
SQ   SEQUENCE   100 AA;  11009 MW;  AB58C605152BB911 CRC64;


  Query Match             10.9%;  Score 70.5;  DB 4;  Length 100;
  Best Local Similarity   29.9%;  Pred. No. 22;
  Matches   23;  Conservative   10;  Mismatches   39;  Indels    5;  Gaps    3;

Qy      21 EMIGPMRPEQFSDEVEPA---TPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQE 77
           |   |:||:|   : :||       |  ::   |   ||: |:| |:|   |    |  |
Db      21 EKTKPLRPQQ-QPQCQPAGGTGQRRGSGSSPSADQQGAQDREEE-AAAAPAPTSRGHRTE 78

Qy      78 QGHPQTGCECEDGPDGQ 94
           :  ||       |  ||
Db      79 KRKPQQPQRRPAGGTGQ 95


Search completed: July  1, 2002, 06:40:24
Job time: 195 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:33:09 ; Search time 23.51 Seconds
(without alignments)
548.046 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 495839

Minimum DB seq length: 0
Maximum DB seq length: 116

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	646	100.0	116	21	AAY83159	GAGE2 polypeptide.
2	638	98.8	116	19	AAW47599	GAGE-2 tumour reje
3	400	61.9	76	21	AAG02123	Human secreted pro
4	271.5	42.0	111	22	AAM39588	Human polypeptide
5	268.5	41.6	112	22	ABG05297	Novel human diagno
6	249.5	38.6	106	22	AAM41374	Human polypeptide
7	229.5	35.5	111	22	AAM78785	Human protein SEQ
8	229.5	35.5	111	22	AAB60500	Human cell cycle a
9	155.5	24.1	60	22	ABG05299	Novel human diagno
10	141	21.8	115	22	AAE08583	Human NOV4 protein
11	120	18.6	102	21	AAY83158	PAGE-4 polypeptide
12	120	18.6	102	21	AAY52998	Human secreted pro
13	113.5	17.6	87	21	AAY83167	PAGE1 polypeptide.
14	109.5	17.0	111	22	AAE08581	Human NOV2 protein
15	108.5	16.8	79	21	AAY83169	PAGE3 polypeptide.
16	104	16.1	89	20	AAY12447	Human 5' EST secre
17	96.5	14.9	111	22	AAE08582	Human NOV3 protein
18	96.5	14.9	111	22	AAE08593	Protein encoded by
19	90	13.9	16	17	AAR99700	Tumour rejection a
20	90	13.9	16	20	AAY29429	GAGE tumour reject
21	89.5	13.9	102	22	ABB29997	Peptide #2648 enco
22	89.5	13.9	102	22	ABB20603	Protein #2602 enco
23	89.5	13.9	102	22	AAM56001	Human brain expres
24	89.5	13.9	102	22	AAM68370	Human bone marrow
25	89.5	13.9	102	22	AAM16187	Peptide #2621 enco
26	89.5	13.9	102	22	AAM28677	Peptide #2714 enco
27	89.5	13.9	102	22	AAM03917	Peptide #2599 enco
28	88	13.6	105	18	AAW32534	Collagen-like poly
29	88	13.6	105	18	AAW12340	Silver halide emul
30	88	13.6	105	19	AAW80239	C-terminal of a Sa
31	88	13.6	105	19	AAW77066	Collagen like poly
32	87	13.5	93	18	AAW32518	Collagen-like poly
33	87	13.5	93	18	AAW12323	Silver halide emul
34	87	13.5	93	19	AAW80223	C-terminal of a Sa
35	87	13.5	93	19	AAW77050	Collagen like poly
36	86	13.3	15	17	AAR97574	Tumour rejection a
37	86	13.3	15	19	AAW41711	Antigenic peptide
38	86	13.3	15	19	AAW41712	Antigenic peptide
39	86	13.3	15	20	AAY29428	GAGE tumour reject
40	86	13.3	107	22	ABB31025	Peptide #3676 enco
41	86	13.3	107	22	ABB36214	Peptide #3720 enco
42	86	13.3	107	22	ABB21591	Protein #3590 enco
43	86	13.3	107	22	AAM56996	Human brain expres
44	86	13.3	107	22	AAM69382	Human bone marrow
45	86	13.3	107	22	AAM17213	Peptide #3647 enco

ALIGNMENTS

RESULT 1
AAY83159
ID AAY83159 standard; Protein; 116 AA.
XX
AC AAY83159;
XX
DT 24-JUL-2000 (first entry)
XX
DE GAGE2 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.
XX
PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues

XX
PS Disclosure; Figure 1a; 63pp; English.
XX
CC PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. The PAGE
CC polypeptide shares sequence similarity with the GAGE and MAGE family
CC of proteins.
XX
SQ Sequence 116 AA;

Query Match 100.0%; Score 646; DB 21; Length 116;
Best Local Similarity 100.0%; Pred. No. 8.3e-53;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
Db 1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy 61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
Db 61 egasagqgpkpeahsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 116

RESULT 2
AAW47599
ID AAW47599 standard; Protein; 116 AA.
XX
AC AAW47599;
XX
DT 30-JUL-1998 (first entry)
XX
DE GAGE-2 tumour rejection antigen precursor.
XX
KW GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW HLA-typing assay.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 38
FT /note= "Ala encoded by GAG"
FT Misc-difference 39
FT /note= "Thr encoded by CCT"
XX
PN WO9749417-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-US10850.
XX
PR 24-JUN-1996; 96US-0669161.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B;
XX
DR WPI; 1998-076905/07.
DR N-PSDB; AAV18717.
XX
PT Isolated nucleic acid encoding GAGE tumour rejection antigen
PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT diagnose melanomas
XX
PS Example 13; Fig 5; 60pp; English.
XX
CC The present sequence represents a GAGE-2 tumour tumour rejection antigen
CC precursor (TRAP). The protein is expressed in a number of tumours. In
CC contrast the only normal tissue which expresses GAGE TRAP protein
CC is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC The major difference between these proteins and GAGE-1 is the absence
CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC TRAP sequence. The rest of the sequences show mismatches at various
CC position, with the exception of GAGE-3 whose 5' end is totally
CC different from the other GAGE cDNAs for the first 112 bases. This
CC region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC The antigens can be used to diagnose melanomas, characterised by
CC expression of a TRAP or presentation of a tumour rejection antigen.
CC Antigens shed into blood or urine can be observed and then used to
CC confirm a diagnosis of melanoma using cytolytic T cell clone
CC proliferation methodologies. Other uses for the processed peptides,
CC include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ Sequence 116 AA;

Query Match 98.8%; Score 638; DB 19; Length 116;
Best Local Similarity 99.1%; Pred. No. 4.6e-52;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
Db 1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy 61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
Db 61 egasagqgpkpeaesqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 116

RESULT 3
AAG02123
ID AAG02123 standard; Protein; 76 AA.
XX
AC AAG02123;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 6204.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.

DR N-PSDB; AAC02129.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 6204; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 76 AA;

Query Match 61.9%; Score 400; DB 21; Length 76;
Best Local Similarity 97.3%; Pred. No. 4e-30;
Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      22 MIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHP 81
           ||||||||||||||||||||||||||| ||||||| ||||||||||||||||||||||
Db       1 migpmrpeqfsdevepatpeegepatqxqdpaaaqxgedegasagqgpkpeahsqeqghp 60

Qy      82 QTGCECEDGPDGQEM 96
           ||||||||||||||
Db      61 qtgcecedgpdgqem 75
```

RESULT 4
AAM39588
ID AAM39588 standard; Protein; 111 AA.
XX
AC AAM39588;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2733.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI58744.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2733; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 111 AA;

Query Match 42.0%; Score 271.5; DB 22; Length 111;
Best Local Similarity 50.0%; Pred. No. 5.4e-18;
Matches 59; Conservative 13; Mismatches 35; Indels 11; Gaps 3;

```
Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQE- 57
           | ||||||||||||| | |||:|||         :||    |:: || |: :|||  ||
Db       1 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpapgqer 52

Qy      58 GEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
            ||:||:  | |  ||  ||    :|| || :||| |    |  |: | || |::| |
Db      53 eedqgaaetqvpdleadlqelsqsktggecgngpddqgkilpkseqfkmpeggdrqpq 110
```

RESULT 5
ABG05297
ID ABG05297 standard; Protein; 112 AA.
XX
AC ABG05297;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #5288.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;

```
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69484.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 35656; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   112 AA;


  Query Match             41.6%;  Score 268.5;  DB 22;  Length 112;
  Best Local Similarity   49.2%;  Pred. No. 1e-17;
  Matches   58;  Conservative   14;  Mismatches   35;  Indels   11;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQE- 57
           | ||||||||||||| | |||:||||        :||    |:: || |: :|||  ||
Db       2 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpapgqer 53

Qy      58 GEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
             ||:|::  | |  ||  ||    :|| || :||| |    |  |: | || |::| |
Db      54 kedqgsaktqvpdleadlqelsqsktggecgngpddqgkilpkseqfkmpeggdrqpq 111


RESULT   6
AAM41374
ID   AAM41374 standard; Protein; 106 AA.
XX
AC   AAM41374;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 6305.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI60530.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 2; SEQ ID NO 6305; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   106 AA;


  Query Match             38.6%;  Score 249.5;  DB 22;  Length 106;
  Best Local Similarity   50.5%;  Pred. No. 5.7e-16;
  Matches   56;  Conservative   10;  Mismatches   34;  Indels   11;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQE- 57
           | ||||||||||||| | |||:||||        :||    |:: || |: :|||  ||
Db       2 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpapgqer 53

Qy      58 GEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPE 108
             ||:||:  | |  ||  ||    :|| || |||| |       |: | ||
Db      54 eedqgaaetqvpdleadlqelsqsktgdecgdgpdvqgkiltkseqfkmpe 104


RESULT   7
AAM78785
ID   AAM78785 standard; Protein; 111 AA.
XX
AC   AAM78785;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1447.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
```

XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51918.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3714-3715; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 111 AA;

Query Match 35.5%; Score 229.5; DB 22; Length 111;
Best Local Similarity 45.8%; Pred. No. 4.3e-14;
Matches 54; Conservative 15; Mismatches 38; Indels 11; Gaps 3;

Qy 1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQEG 58
Db 1 mswrgrstyrprprrslqppeligam--------leptdeepkeekpptksrnptpdqkr 52

Qy 59 E-DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
Db 53 eddqgaaeiqvpdleadlqelcqtktgdgceggtdvkgkilpkaehfkmpeagegksq 110

RESULT 8
AAB60500
ID AAB60500 standard; Protein; 111 AA.
XX
AC AAB60500;
XX
DT 24-APR-2001 (first entry)
XX
DE Human cell cycle and proliferation protein CCYPR-48, SEQ ID NO:48.
XX
KW Cell cycle and proliferation protein; CCYPR; human; agonist;
KW antagonist; gene therapy; detection; gene therapy;
KW transgenic animal disease model; immune disorder;
KW developmental disorder; cell signalling disorder;
KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
KW arteriosclerosis; asthma; allergy; diabetes mellitus;
KW menstrual cycle disorder; bacterial infection.
XX
OS Homo sapiens.
XX
PN WO200107471-A2.
XX
PD 01-FEB-2001.
XX
PF 21-JUL-2000; 2000WO-US19948.
XX
PR 21-JUL-1999; 99US-0145075.
PR 08-SEP-1999; 99US-0153129.
PR 10-NOV-1999; 99US-0164647.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;
XX
DR WPI; 2001-112727/12.
DR N-PSDB; AAF59637.
XX
PT Human cell cycle and proliferation proteins and polynucleotides are
PT used to treat, diagnose and prevent immune, developmental and cell
PT signaling disorders and cell proliferative disorders including cancer -
XX
PS Claim 1; Page 158; 205pp; English.
XX
CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and
CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions
CC associated with decreased expression of functional CCYPR, while CCYPR
CC antagonists are used to treat diseases or conditions associated with
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or
CC radioimmunoasays to detect CCYPR. CCYPR itself may be used to detect
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)
CC that specifically bind to CCYPR, and in drug screening methods to
CC identify compounds that modulate the activity of CCYPR. CCYPR
CC nucleotides can be used to generate transgenic animal models of human
CC disease, and can be used in gene therapy in target cells with genetic
CC abnormalities with respect to the expression of CCYPR for the
CC treatment or prevention of a disorder associated with CCYPR.
CC Diseases which can be diagnosed, treated and prevented using CCYPR
CC proteins, nucleic acids, agonists or antagonists include immune,
CC developmental and cell signalling disorders, and cell proliferative
CC disorders including cancer. Specific examples of these disorders
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,
CC diabetes mellitus, disorders of the menstrual cycle and infections
CC caused by bacteria.
XX
SQ Sequence 111 AA;

Query Match 35.5%; Score 229.5; DB 22; Length 111;
Best Local Similarity 45.8%; Pred. No. 4.3e-14;
Matches 54; Conservative 15; Mismatches 38; Indels 11; Gaps 3;

Qy 1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQEG 58
Db 1 mswrgrstyrprprrslqppeligam--------leptdeepkeekpptksrnptpdqkr 52

Qy 59 E-DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
Db 53 eddqgaaeiqvpdleadlqelcqtktgdgceggtdvkgkilpkaehfkmpeagegksq 110

RESULT 9
ABG05299

```
ID   ABG05299 standard; Protein; 60 AA.
XX
AC   ABG05299;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5290.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69486.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 35658; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   60 AA;


  Query Match             24.1%;  Score 155.5;  DB 22;  Length 60;
  Best Local Similarity   50.0%;  Pred. No. 1.7e-07;
  Matches   34;  Conservative    7;  Mismatches   14;  Indels   13;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQEG 58
           | ||||||||||||| | |||:|||         :||    |:: || |: :|||    |
Db       2 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpa---pg 50

Qy      59 EDEGASAG 66
           |   :|
Db      51 sgerrrsg 58
```

```
RESULT  10
AAE08583
ID   AAE08583 standard; Protein; 115 AA.
XX
AC   AAE08583;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Human NOV4 protein.
XX
KW   Human; NOVX; G-antigen; GAGE-like protein; interferon;
KW   G-protein coupled receptor; GPCR; hepatocyte nuclear factor;
KW   mast cell protease; gene therapy; proliferative disorder; cancer;
KW   immune disorder; hepatic disorder; cirrhosis; viral infection;
KW   hepatitis; neuroolfactory system-related disorder; neurological disorder;
KW   Parkinson's disease; infertility; autoimmune disease; arthritis;
KW   multiple sclerosis; allergy; wound healing; cytostatic; nootropic;
KW   immunosuppressive; neuroprotective; vulnerary; hepatotropic.
XX
OS   Homo sapiens.
XX
PN   WO200161009-A2.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US04828.
XX
PR   15-FEB-2000; 2000US-0182723.
PR   15-FEB-2000; 2000US-0182724.
PR   15-FEB-2000; 2000US-0182733.
PR   22-FEB-2000; 2000US-0183896.
PR   23-FEB-2000; 2000US-0184275.
PR   23-FEB-2000; 2000US-0184482.
PR   23-FEB-2000; 2000US-0184497.
PR   24-FEB-2000; 2000US-0184744.
PR   13-APR-2000; 2000US-0197083.
PR   10-AUG-2000; 2000US-0224157.
PR   18-SEP-2000; 2000US-0233405.
PR   27-SEP-2000; 2000US-0236060.
PR   02-JAN-2001; 2001US-0259414.
PR   18-JAN-2001; 2001US-0262454.
PR   14-FEB-2001; 2001US-0783429.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Malyankar UM,  Tchernev VT,  Padigaru M,  Taupier RJ,  Spytek KA;
PI   Majumder K,  Guo X,  Spaderna SK,  Boldog FL;
XX
DR   WPI; 2001-514775/56.
DR   N-PSDB; AAD14983.
XX
PT   Isolated novel polypeptides useful for diagnosis of and treating
PT   cancer, infertility, autoimmune diseases, arthritis, multiple
PT   sclerosis, allergies, wound healing and hepatic disorders -
XX
PS   Claim 1; Page 14; 140pp; English.
XX
CC   The present sequence is a human NOV4 protein. The NOVX protein has
CC   homology with one of G-antigen (GAGE)-like protein, interferon,
CC   G-protein coupled receptor (GPCR), hepatocyte nuclear factor or mast
CC   cell protease. The NOVX is useful for treating or preventing a pathology
CC   associated with NOVX. It is also useful for determining the presence or
CC   amount of NOVX DNA in a sample, for identifying a potential therapeutic
CC   agent and in gene therapy. It is also useful for determining the presence
CC   of or predisposition to a disease associated with altered levels of NOVX.
CC   It is also useful for the diagnosis and treatment of proliferative
CC   disorders, e.g., cancer, immune disorders, hepatic disorders, e.g.,
CC   cirrhosis, viral infections, e.g., hepatitis, neuroolfactory
CC   system-related disorders, neurological disorders, e.g., Parkinson's
CC   disease, infertility, autoimmune diseases, arthritis, multiple sclerosis,
CC   allergies and wound healing.
XX
```

```
SQ   Sequence   115 AA;

  Query Match             21.8%;  Score 141;  DB 22;  Length 115;
  Best Local Similarity   51.9%;  Pred. No. 7.6e-06;
  Matches   28;  Conservative    9;  Mismatches    7;  Indels   10;  Gaps    2;

Qy      1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDP 52
          |||||||||||||||| ::||||:|| |        :||    |:| :| |: ::|
Db      1 mswrgrstyrprprrslqppeligam--------leptdeepkeekpptksrnp 46


RESULT  11
AAY83158
ID   AAY83158 standard; Protein; 102 AA.
XX
AC   AAY83158;
XX
DT   24-JUL-2000  (first entry)
XX
DE   PAGE-4 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
DR   N-PSDB; AAZ93524.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems.
XX
SQ   Sequence   102 AA;

  Query Match             18.6%;  Score 120;  DB 21;  Length 102;
  Best Local Similarity   34.5%;  Pred. No. 0.0006;
  Matches   39;  Conservative   10;  Mismatches   52;  Indels   12;  Gaps    3;

Qy      1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
          || | ||  | |     | |:::           | |   :: || |  ||    |  |
Db      1 msarvrsrsrgrgdgq-eapdvva---------fvapgesqqeepptdnqdiepgq--er 48

Qy     61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
          ||    :  | |   ||    :|  |  || | :|  ||||:  || | |: |
Db     49 egtppieerkvegdcqemdlektrsergdgsdvkektppnpkhaktkeagdgq 101


RESULT  12
AAY52998
ID   AAY52998 standard; Protein; 102 AA.
XX
AC   AAY52998;
XX
DT   29-FEB-2000  (first entry)
XX
DE   Human secreted protein clone bn365_53 protein sequence SEQ ID NO:2.
XX
KW   Human; secreted protein; nutritional; cytokine; cell proliferation;
KW   differentiation; immune stimulating; vaccine; suppression;
KW   haematopoiesis regulation; tissue growth; activin; inhibin;
KW   chemotactic; chemokinetic; haemostatic; thrombolytic; receptor;
KW   ligand; anti-inflammatory; cadherin; tumour invasion suppressor;
KW   tumour inhibition; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO9957132-A1.
XX
PD   11-NOV-1999.
XX
PF   07-MAY-1999;   99WO-US09970.
XX
PR   07-MAY-1998;   98US-0084564.
PR   02-JUN-1998;   98US-0087645.
PR   22-JUL-1998;   98US-0093712.
PR   31-JUL-1998;   98US-0094935.
PR   10-AUG-1998;   98US-0095880.
PR   11-AUG-1998;   98US-0096068.
PR   06-MAY-1999;   99US-0096068.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Agostino MJ,  Steininger RJ,  Bowman MR;
PI   DiBlasio-Smith E,  Widom A;
XX
DR   WPI; 2000-052937/04.
DR   N-PSDB; AAZ33316.
XX
PT   New polynucleotides encoding secreted human proteins, derived from
PT   adult placenta, adult retina, fetal brain, fetal -
XX
PS   Claim 9; Page 360-361; 492pp; English.
XX
CC   The present invention describes new human secreted proteins which were
CC   isolated from adult placenta, adult retina, foetal brain, foetal kidney,
CC   adult blood, adult brain, adult thyroid, adult bladder, adult neural
CC   tissue, adult testes, and adult lymph node cDNA libraries. The human
CC   secreted proteins, and the polynucleotides encoding them, are predicted
CC   to have biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals. Suggested activities include nutritional activity, cytokine
CC   and cell proliferation/differentiation activity, immune stimulating
CC   (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
```

```
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, and tumour inhibition
CC   activity. The polynucleotides are also stated to be useful for gene
CC   therapy. AAZ33316 to AAZ33373 encode human secreted proteins, and
CC   AAY52998 to AAY53060 represent human secreted proteins, given in the
CC   present invention.
XX
SQ   Sequence   102 AA;


  Query Match             18.6%;  Score 120;  DB 21;  Length 102;
  Best Local Similarity   34.5%;  Pred. No. 0.0006;
  Matches   39;  Conservative   10;  Mismatches   52;  Indels   12;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           || | ||  | |     | |:::          | |   :: || |  ||    |  |
Db       1 msarvrsrsrgrgdgq-eapdvva---------fvapgesqqeepptdnqdiepgq--er 48

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           ||     :  | |   ||    :|  |  || | :|  ||||:  || | |: |
Db      49 egtppieerkvegdcqemdlektrsergdgsdvkektppnpkhaktkeagdgq 101


RESULT  13
AAY83167
ID   AAY83167 standard; Protein; 87 AA.
XX
AC   AAY83167;
XX
DT   24-JUL-2000  (first entry)
XX
DE   PAGE1 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1b; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. This sequence is
CC   of the PAGE1 polypeptide which shares sequence similarity with
CC   PAGE-4.
XX
SQ   Sequence   87 AA;


  Query Match             17.6%;  Score 113.5;  DB 21;  Length 87;
  Best Local Similarity   32.7%;  Pred. No. 0.0021;
  Matches   37;  Conservative    8;  Mismatches   41;  Indels   27;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           || | ||  | |     | |:::          | |   :| || |
Db       1 msarvrsrsrgrgdgq-eapdvva---------fvapgesqeeepptd------------ 38

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
                |||  ||  ||    :|  |  || | :|  ||||:  || | |: |
Db      39 -----nqgpdmeafqqeldlektrsergdgsdvkektppnpkhaktkeagdgq 86


RESULT  14
AAE08581
ID   AAE08581 standard; Protein; 111 AA.
XX
AC   AAE08581;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Human NOV2 protein.
XX
KW   Human; NOVX; G-antigen; GAGE-like protein; interferon;
KW   G-protein coupled receptor; GPCR; hepatocyte nuclear factor;
KW   mast cell protease; gene therapy; proliferative disorder; cancer;
KW   immune disorder; hepatic disorder; cirrhosis; viral infection;
KW   hepatitis; neuroolfactory system-related disorder; neurological disorder;
KW   Parkinson's disease; infertility; autoimmune disease; arthritis;
KW   multiple sclerosis; allergy; wound healing; cytostatic; nootropic;
KW   immunosuppressive; neuroprotective; vulnerary; hepatotropic.
XX
OS   Homo sapiens.
XX
PN   WO200161009-A2.
XX
PD   23-AUG-2001.
XX
PF   15-FEB-2001; 2001WO-US04828.
XX
PR   15-FEB-2000; 2000US-0182723.
PR   15-FEB-2000; 2000US-0182724.
PR   15-FEB-2000; 2000US-0182733.
PR   22-FEB-2000; 2000US-0183896.
PR   23-FEB-2000; 2000US-0184275.
PR   23-FEB-2000; 2000US-0184482.
PR   23-FEB-2000; 2000US-0184497.
PR   24-FEB-2000; 2000US-0184744.
PR   13-APR-2000; 2000US-0197083.
PR   10-AUG-2000; 2000US-0224157.
PR   18-SEP-2000; 2000US-0233405.
PR   27-SEP-2000; 2000US-0236060.
PR   02-JAN-2001; 2001US-0259414.
PR   18-JAN-2001; 2001US-0262454.
PR   14-FEB-2001; 2001US-0783429.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Malyankar UM,  Tchernev VT,  Padigaru M,  Taupier RJ,  Spytek KA;
PI   Majumder K,  Guo X,  Spaderna SK,  Boldog FL;
```

XX
DR WPI; 2001-514775/56.
DR N-PSDB; AAD14981.
XX
PT Isolated novel polypeptides useful for diagnosis of and treating
PT cancer, infertility, autoimmune diseases, arthritis, multiple
PT sclerosis, allergies, wound healing and hepatic disorders -
XX
PS Claim 1; Page 9; 140pp; English.
XX
CC The present sequence is a human NOV2 protein. The NOVX protein has
CC homology with one of G-antigen (GAGE)-like protein, interferon,
CC G-protein coupled receptor (GPCR), hepatocyte nuclear factor or mast
CC cell protease. The NOVX is useful for treating or preventing a pathology
CC associated with NOVX. It is also useful for determining the presence or
CC amount of NOVX DNA in a sample, for identifying a potential therapeutic
CC agent and in gene therapy. It is also useful for determining the presence
CC of or predisposition to a disease associated with altered levels of NOVX.
CC It is also useful for the diagnosis and treatment of proliferative
CC disorders, e.g., cancer, immune disorders, hepatic disorders, e.g.,
CC cirrhosis, viral infections, e.g., hepatitis, neuroolfactory
CC system-related disorders, neurological disorders, e.g., Parkinson's
CC disease, infertility, autoimmune diseases, arthritis, multiple sclerosis,
CC allergies and wound healing.
XX
SQ Sequence 111 AA;

Query Match 17.0%; Score 109.5; DB 22; Length 111;
Best Local Similarity 32.7%; Pred. No. 0.0063;
Matches 36; Conservative 17; Mismatches 50; Indels 7; Gaps 2;

```
Qy       4 RGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGA 63
           | :|: |   :   :|   :| :  :: ::|      :| || |  |  | : | |:|||
Db       8 rsqssergndqessqp---vgsvivqepteekr----qeeepptdnqgiapsgeienega 60

Qy      64 SAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
             | |||  ||  ||    :   |  ||||:|   |  :  |  | |: |
Db      61 pavqgpdmeafqqelallkiedepgdgpdvregimptfdltkvleagdaq 110
```

RESULT 15
AAY83169
ID AAY83169 standard; Protein; 79 AA.
XX
AC AAY83169;
XX
DT 24-JUL-2000 (first entry)
XX
DE PAGE3 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.
XX
PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues
XX
PS Disclosure; Figure 1b; 63pp; English.
XX
CC PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. This sequence is
CC of the PAGE3 polypeptide which shares sequence similarity with
CC PAGE-4.
XX
SQ Sequence 79 AA;

Query Match 16.8%; Score 108.5; DB 21; Length 79;
Best Local Similarity 43.5%; Pred. No. 0.0054;
Matches 30; Conservative 5; Mismatches 33; Indels 1; Gaps 1;

```
Qy      45 PATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEV 104
           |  : ||    ||  ||||   | |   |:  |   |:|| |  |||: :    || | |
Db       9 ppiesqdytpgqe-rdegaldfqvpslaaylweltrpktggergdgpnvkgeslpnlepv 67

Qy     105 KTPEEGEKQ 113
           | || || |
Db      68 kipeagegq 76
```

Search completed: July 1, 2002, 06:36:35
Job time: 206 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:11:45 ; Search time 1773.93 Seconds
(without alignments)
6346.630 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	538	100.0	538	6	AR028488	AR028488 Sequence
2	538	100.0	538	6	I55851	I55851 Sequence 14
3	533.2	99.1	551	9	BC018052	BC018052 Homo sapi
4	531.6	98.8	1245	6	AX285022	AX285022 Sequence
5	530	98.5	530	9	HSU19143	U19143 Human GAGE-
6	525.4	97.7	528	9	AF055473	AF055473 Homo sapi
7	512.8	95.3	540	6	AR028490	AR028490 Sequence
8	512.8	95.3	540	6	I55853	I55853 Sequence 16
9	510.2	94.8	539	6	AR028492	AR028492 Sequence
10	510.2	94.8	539	6	I55855	I55855 Sequence 18
11	504.8	93.8	532	6	AR028491	AR028491 Sequence
12	504.8	93.8	532	6	I55854	I55854 Sequence 17
13	501.4	93.2	528	6	AX334151	AX334151 Sequence
14	501.4	93.2	528	9	HSU19145	U19145 Human GAGE-
15	498.8	92.7	527	9	HSU19147	U19147 Human GAGE-
16	497.6	92.5	524	9	AF058988	AF058988 Homo sapi
17	496.8	92.3	524	9	HSU19146	U19146 Human GAGE-
18	488	90.7	526	9	AF055474	AF055474 Homo sapi
19	429.2	79.8	560	6	AR028489	AR028489 Sequence
20	429.2	79.8	560	6	I55852	I55852 Sequence 15
21	421.2	78.3	552	9	HSU19144	U19144 Human GAGE-
22	374.4	69.6	646	6	AR028482	AR028482 Sequence
23	374.4	69.6	646	6	I55845	I55845 Sequence 1
24	374.4	69.6	646	9	HSU19142	U19142 Human GAGE-
25	374.4	69.6	648	6	I57317	I57317 Sequence 1
c 26	301.4	56.0	530	6	AX284300	AX284300 Sequence
c 27	299.2	55.6	365	6	AX284692	AX284692 Sequence
28	209.6	39.0	662	9	BC004861	BC004861 Homo sapi
29	209.6	39.0	676	9	AF058989	AF058989 Homo sapi
30	185.8	34.5	648	9	BC009232	BC009232 Homo sapi
31	182.4	33.9	493	9	HSA318881	AJ318881 Homo sapi
32	179.8	33.4	620	9	HSA318880	AJ318880 Homo sapi
33	173.8	32.3	611	6	AX226501	AX226501 Sequence
c 34	163.2	30.3	79539	9	AC093664	AC093664 Homo sapi
c 35	163.2	30.3	240000	2	AC009528	AC009528 Homo sapi
36	156	29.0	580	6	AX078298	AX078298 Sequence
37	146.8	27.3	763	9	BC009230	BC009230 Homo sapi
c 38	140.8	26.2	20587	9	HSL185E6A	Z68274 Human DNA s
c 39	140.8	26.2	62493	9	HS193G15	AL117391 Human DNA
c 40	140.8	26.2	62493	9	HSJ193G15	AL096838 Human DNA
c 41	140.8	26.2	185713	2	AL645949	AL645949 Homo sapi
42	131.4	24.4	9531	9	AF055475	AF055475 Homo sapi
43	131.4	24.4	48802	2	AC068431	AC068431 Homo sapi
44	130.4	24.2	475	6	AX226497	AX226497 Sequence
45	129.8	24.1	194418	9	AF235098	AF235098 Homo sapi

ALIGNMENTS

RESULT 1
AR028488
LOCUS AR028488 538 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 14 from patent US 5858689.
ACCESSION AR028488
VERSION AR028488.1 GI:5940461
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 538)
AUTHORS van der Bruggen,P., van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
TITLE Isolated peptides derived from the gage tumor rejection antigen precursor and uses thereof
JOURNAL Patent: US 5858689-A 14 12-JAN-1999;
FEATURES Location/Qualifiers
source 1. .538
/organism="unknown"
BASE COUNT 160 a 116 c 155 g 107 t
ORIGIN

```
  Query Match           100.0%;  Score 538;  DB 6;  Length 538;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-120;
  Matches  538;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538


RESULT   2
I55851
LOCUS       I55851                 538 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 14 from patent US 5648226.
ACCESSION   I55851
VERSION     I55851.1  GI:2476645
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 538)
  AUTHORS   Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
  TITLE     Isolated peptides derived from tumor rejection antigens, and their
            use
  JOURNAL   Patent: US 5648226-A 14 15-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .538
                     /organism="unknown"
BASE COUNT      160 a    116 c    155 g    107 t
ORIGIN


  Query Match           100.0%;  Score 538;  DB 6;  Length 538;
  Best Local Similarity  100.0%;  Pred. No. 6.6e-120;
  Matches  538;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538


RESULT   3
BC018052
LOCUS       BC018052               551 bp    mRNA    linear   PRI 06-DEC-2001
DEFINITION  Homo sapiens, G antigen 8, clone MGC:26395 IMAGE:4812462, mRNA,
            complete cds.
ACCESSION   BC018052
VERSION     BC018052.1  GI:17390105
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 32 Row: k Column: 1
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503878.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Homo sapiens"
                     /db_xref="LocusID:26749"
```

```
                /db_xref="taxon:9606"
                /clone="MGC:26395 IMAGE:4812462"
                /tissue_type="Brain, hippocampus"
                /clone_lib="NIH_MGC_95"
                /lab_host="DH10B"
                /note="Vector: pBluescript"
     CDS        97. .447
                /codon_start=1
                /product="G antigen 8"
                /protein_id="AAH18052.1"
                /db_xref="GI:17390106"
                /translation="MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGE
                PATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPE
                EVKTPEEGEKQSQC"
BASE COUNT      164 a    118 c    160 g    109 t
ORIGIN
```

```
  Query Match           99.1%;  Score 533.2;  DB 9;  Length 551;
  Best Local Similarity  99.4%;  Pred. No. 9.5e-119;
  Matches  535;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     14 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 73

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     74 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 133

Qy    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    134 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 193

Qy    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    194 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 253

Qy    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    254 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 313

Qy    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    314 CTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 373

Qy    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    374 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 433

Qy    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    434 AATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 493

Qy    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    494 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 551
```

```
RESULT   4
AX285022
LOCUS       AX285022                1245 bp    DNA     linear   PAT 20-NOV-2001
DEFINITION  Sequence 827 from Patent WO0179556.
ACCESSION   AX285022
VERSION     AX285022.1  GI:17045710
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Lillie,J., Brown,J.L., Bolt,A. and van Huffel,C.
  TITLE     Novel genes, compositions and methods for the identification,
            assessment, prevention, and therapy of human cancers
  JOURNAL   Patent: WO 0179556-A 827 25-OCT-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1245
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      382 a    308 c    390 g    153 t     12 others
ORIGIN
```

```
  Query Match           98.8%;  Score 531.6;  DB 6;  Length 1245;
  Best Local Similarity  99.3%;  Pred. No. 2.4e-118;
  Matches  534;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    179 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 238

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    239 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTGGCCTAGAC 298

Qy    121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    299 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 358

Qy    181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 418

Qy    241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    419 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 478

Qy    301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    479 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 538

Qy    361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 598

Qy    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    599 AATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 658

Qy    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||||| |||||||||| |
Db    659 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAGA 716
```

```
RESULT   5
HSU19143
LOCUS       HSU19143                530 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-2 protein mRNA, complete cds.
ACCESSION   U19143
VERSION     U19143.1  GI:914900
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 530)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (28-DEC-1994) Benoit J Van Den Eynde, Ludwig Institute
            For Cancer Research, 74 Avenue Hippocrate, BRUSSELS, 1200, BELGIUM
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             84. .434
                     /codon_start=1
                     /product="GAGE-2 protein"
                     /protein_id="AAA82745.1"
                     /db_xref="GI:914901"
                     /translation="MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGE
                     PATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPE
                     EVKTPEEGEKQSQC"
BASE COUNT      152 a    116 c    155 g    107 t
ORIGIN


  Query Match             98.5%;  Score 530;  DB 9;  Length 530;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-118;
  Matches  530;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530


RESULT   6
AF055473
LOCUS       AF055473                 528 bp    mRNA    linear   PRI 01-MAY-2000
DEFINITION  Homo sapiens GAGE-8 mRNA, complete cds.
ACCESSION   AF055473
VERSION     AF055473.1  GI:3511022
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   De Backer,O., Arden,K.C., Boretti,M., Vantomme,V., De Smet,C.,
            Czekay,S., Viars,C.S., De Plaen,E., Brasseur,F., Chomez,P., Van den
            Eynde,B., Boon,T. and van der Bruggen,P.
  TITLE     Characterization of the GAGE genes that are expressed in various
            human cancers and in normal testis
  JOURNAL   Cancer Res. 59 (13), 3157-3165 (1999)
  MEDLINE   99323388
   PUBMED   10397259
REFERENCE   2  (bases 1 to 528)
  AUTHORS   De Backer,O.R.Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1998) Ludwig Institute for Cancer Research,
            Brussels Branch, 74 av. Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             73. .423
                     /codon_start=1
                     /product="GAGE-8"
                     /protein_id="AAC33676.1"
                     /db_xref="GI:3511023"
                     /translation="MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGE
                     PATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPE
                     EVKTPEEGEKQSQC"
BASE COUNT      158 a    112 c    151 g    107 t
ORIGIN


  Query Match             97.7%;  Score 525.4;  DB 9;  Length 528;
  Best Local Similarity   99.8%;  Pred. No. 7.3e-117;
  Matches  526;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      12 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 71
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 60

Qy      72 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTAC 131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACGCTAC 120

Qy     132 GTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAA 191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAA 180

Qy     192 CCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAG 251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAG 240

Qy     252 GAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAG 311
           |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     241 GAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAG 300

Qy     312 GAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATG 371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATG 360

Qy     372 GACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGT 431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGT 420

Qy     432 TAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAA 491
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAA 480

Qy     492 ATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 527


RESULT   7
```

```
AR028490
LOCUS       AR028490                 540 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 16 from patent US 5858689.
ACCESSION   AR028490
VERSION     AR028490.1  GI:5940463
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   van der Bruggen,P., van den Eynde,B., DeBacker,O. and
            Boon-Falleur,T.
  TITLE     Isolated peptides derived from the gage tumor rejection antigen
            precursor and uses thereof
  JOURNAL   Patent: US 5858689-A 16 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="unknown"
BASE COUNT      159 a    114 c    156 g    111 t
ORIGIN
```

```
  Query Match             95.3%;  Score 512.8;  DB 6;  Length 540;
  Best Local Similarity   98.1%;  Pred. No. 8.2e-114;
  Matches  530;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy       2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy      62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db      61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy     119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
           |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180

Qy     179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy     239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy     299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy     359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420

Qy     419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy     479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 540
```

```
RESULT   8
I55853
LOCUS       I55853                   540 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 16 from patent US 5648226.
ACCESSION   I55853
VERSION     I55853.1  GI:2476647
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
  TITLE     Isolated peptides derived from tumor rejection antigens, and their
            use
  JOURNAL   Patent: US 5648226-A 16 15-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .540
                     /organism="unknown"
BASE COUNT      159 a    114 c    156 g    111 t
ORIGIN
```

```
  Query Match             95.3%;  Score 512.8;  DB 6;  Length 540;
  Best Local Similarity   98.1%;  Pred. No. 8.2e-114;
  Matches  530;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy       2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy      62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
           ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db      61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy     119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
           |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180

Qy     179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy     239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy     299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
           |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy     359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420

Qy     419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy     479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 540
```

```
RESULT   9
AR028492
LOCUS       AR028492                 539 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 18 from patent US 5858689.
ACCESSION   AR028492
VERSION     AR028492.1  GI:5940465
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   van der Bruggen,P., van den Eynde,B., DeBacker,O. and
            Boon-Falleur,T.
  TITLE     Isolated peptides derived from the gage tumor rejection antigen
            precursor and uses thereof
  JOURNAL   Patent: US 5858689-A 18 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="unknown"
BASE COUNT      158 a    113 c    157 g    111 t
```

```
ORIGIN

  Query Match             94.8%;  Score 510.2;  DB 6;  Length 539;
  Best Local Similarity   98.0%;  Pred. No. 3.5e-113;
  Matches  528;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 60

Qy      63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
           |||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||
Db      61 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGA 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||| ||||| ||| ||||||||||| ||||||||||||||||||||||||||||||||
Db     121 CCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGGCAGGAGGTGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 539


RESULT  10
I55855
LOCUS       I55855                   539 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 18 from patent US 5648226.
ACCESSION   I55855
VERSION     I55855.1  GI:2476649
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
  TITLE     Isolated peptides derived from tumor rejection antigens, and their
            use
  JOURNAL   Patent: US 5648226-A 18 15-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="unknown"
BASE COUNT      158 a    113 c    157 g    111 t
ORIGIN


  Query Match             94.8%;  Score 510.2;  DB 6;  Length 539;
  Best Local Similarity   98.0%;  Pred. No. 3.5e-113;
  Matches  528;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 60

Qy      63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
           |||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||
Db      61 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGA 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||| ||||| ||| ||||||||||| ||||||||||||||||||||||||||||||||
Db     121 CCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGGCAGGAGGTGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAAA 539


RESULT  11
AR028491
LOCUS       AR028491                 532 bp    DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 17 from patent US 5858689.
ACCESSION   AR028491
VERSION     AR028491.1  GI:5940464
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   van der Bruggen,P., van den Eynde,B., DeBacker,O. and
            Boon-Falleur,T.
  TITLE     Isolated peptides derived from the gage tumor rejection antigen
            precursor and uses thereof
  JOURNAL   Patent: US 5858689-A 17 12-JAN-1999;
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="unknown"
BASE COUNT      156 a    111 c    154 g    111 t
ORIGIN


  Query Match             93.8%;  Score 504.8;  DB 6;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 7.1e-112;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           ||||||||||||||||||||||||||||||||||||||   ||| ||||||||||||| |
Db      61 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGC 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     121 GCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 180
```

```
Qy    187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 240

Qy    247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    241 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATA 300

Qy    307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 360

Qy    367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 420

Qy    427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 480

Qy    487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 532
```

RESULT 12
I55854

```
LOCUS       I55854                   532 bp    DNA     linear   PAT 07-OCT-1997
DEFINITION  Sequence 17 from patent US 5648226.
ACCESSION   I55854
VERSION     I55854.1  GI:2476648
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Van den Eynde,B., DeBacker,O. and Boon-Falleur,T.
  TITLE     Isolated peptides derived from tumor rejection antigens, and their
            use
  JOURNAL   Patent: US 5648226-A 17 15-JUL-1997;
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="unknown"
BASE COUNT      156 a    111 c    154 g    111 t
ORIGIN
```

```
  Query Match             93.8%;  Score 504.8;  DB 6;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 7.1e-112;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy     10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 60

Qy     70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
          |||||||||||||||||||||||||||||||||||||    ||| ||||||||||||| |
Db     61 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGC 120

Qy    127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
          |||| ||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 180

Qy    187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 240

Qy    247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    241 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATA 300

Qy    307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 360

Qy    367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 420

Qy    427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 480

Qy    487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 532
```

RESULT 13
AX334151

```
LOCUS       AX334151                 528 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 4660 from Patent WO0194629.
ACCESSION   AX334151
VERSION     AX334151.1  GI:18124870
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 4660 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a    114 c    156 g    111 t
ORIGIN
```

```
  Query Match             93.2%;  Score 501.4;  DB 6;  Length 528;
  Best Local Similarity   98.3%;  Pred. No. 4.7e-111;
  Matches  518;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy     62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
          |||||||||||||||||||||||||||||||||||||||||||||    ||| |||||||
Db     61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
          |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420
```

```
Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy    479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 527


RESULT 14
HSU19145
LOCUS       HSU19145                 528 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-4 protein mRNA, complete cds.
ACCESSION   U19145
VERSION     U19145.1  GI:914904
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 528)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1994) BENOIT J VAN DEN EYNDE, Ludwig Institute
            for Cancer Research, 74 Avenue Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             83. .436
                     /codon_start=1
                     /product="GAGE-4 protein"
                     /protein_id="AAA82747.1"
                     /db_xref="GI:914905"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      147 a    114 c    156 g    111 t
ORIGIN


  Query Match             93.2%;  Score 501.4;  DB 9;  Length 528;
  Best Local Similarity   98.3%;  Pred. No. 4.7e-111;
  Matches  518;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 60

Qy     62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
          ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db     61 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAG 120

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
          |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 180

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 420

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          ||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    421 GCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 480

Qy    479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 527


RESULT 15
HSU19147
LOCUS       HSU19147                 527 bp    mRNA    linear   PRI 04-DEC-1995
DEFINITION  Human GAGE-6 protein mRNA, complete cds.
ACCESSION   U19147
VERSION     U19147.1  GI:914908
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S.
            and Boon,T.
  TITLE     A new family of genes coding for an antigen recognized by
            autologous cytolytic T lymphocytes on a human melanoma
  JOURNAL   J. Exp. Med. 182 (3), 689-698 (1995)
  MEDLINE   95378788
REFERENCE   2  (bases 1 to 527)
  AUTHORS   Van Den Eynde,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1994) BENOIT J VAN DEN EYNDE, Ludwig Institute
            for Cancer Research, 74 Avenue Hippocrate, Brussels B-1200, Belgium
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /sex="female"
                     /cell_line="MZ2-MEL.43"
                     /tissue_type="melanoma"
                     /dev_stage="adult"
     CDS             82. .435
                     /codon_start=1
                     /product="GAGE-6 protein"
                     /protein_id="AAA82749.1"
                     /db_xref="GI:914909"
                     /translation="MSWRGRSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEG
                     EPATQRQDPAAAQEGEDEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEVDPPNP
                     EEVKTPEEGEKQSQC"
BASE COUNT      146 a    113 c    157 g    111 t
ORIGIN


  Query Match             92.7%;  Score 498.8;  DB 9;  Length 527;
  Best Local Similarity   98.1%;  Pred. No. 2e-110;
  Matches  516;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 60

Qy     63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
          ||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||
```

```
Db      61 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGA 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||| ||||| ||| |||||||||||| ||||||||||||||||||||||||||||||
Db     121 CCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGGCAGGAGGTGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 525
           |||||||||||||||||||||||||||||||||||||||||||||
Db     481 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAA 526
```

Search completed: July 1, 2002, 07:13:58
Job time: 3733 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:11:45 ; Search time 1625.58 Seconds
(without alignments)
4466.936 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	524.6	97.5	569	10	BG700165	BG700165 602679622
	2	492.6	91.6	813	10	BG773070	BG773070 602721370
	3	492	91.4	527	10	BI826605	BI826605 603077056
	4	492	91.4	851	10	BG250953	BG250953 602363801
	5	490.6	91.2	572	10	BG120838	BG120838 602351010
	6	476.8	88.6	509	10	BI868671	BI868671 603392594
	7	463.4	86.1	623	10	BM172302	BM172302 imageqc_4
c	8	452.4	84.1	489	9	AA447559	AA447559 zw81e11.s
c	9	431.4	80.2	457	9	AW510753	AW510753 hd39d05.x
c	10	424.6	78.9	447	9	AI381509	AI381509 te76b07.x
c	11	424.4	78.9	464	9	AA738037	AA738037 nx15e11.s
	12	424	78.8	581	10	BG700088	BG700088 602679431
c	13	413.8	76.9	418	9	AW016546	AW016546 UI-H-BI0p
	14	413.4	76.8	480	10	BG120336	BG120336 602353732
c	15	412.8	76.7	426	9	AA868226	AA868226 ak48h07.s
c	16	410.8	76.4	450	9	AA760996	AA760996 nx32h08.s
c	17	401.2	74.6	455	9	AI187350	AI187350 qf29a05.x
	18	370.4	68.8	795	10	BI560935	BI560935 603254321
c	19	365.4	67.9	412	9	AW102587	AW102587 xd67g07.x
	20	360.4	67.0	445	11	BC005363	BC005363 Homo sapi
	21	359.4	66.8	419	9	AA448542	AA448542 zw81e11.r
	22	346.6	64.4	839	10	BG527731	BG527731 602556670
c	23	329.8	61.3	384	9	AA913206	AA913206 om53f03.s
	24	318.8	59.3	397	10	BG206349	BG206349 RST25797
	25	309.8	57.6	1069	10	BE777698	BE777698 601348269
c	26	296.4	55.1	398	9	AA918604	AA918604 ol53f02.s
c	27	250.4	46.5	256	10	BG181480	BG181480 RST328 At
	28	233.4	43.4	258	10	BG186708	BG186708 RST5683 A
	29	228.2	42.4	253	10	BG184057	BG184057 RST2973 A
	30	226.6	42.1	261	10	BG199060	BG199060 RST18338
	31	224.8	41.8	245	10	BG212621	BG212621 RST32216
	32	219.8	40.9	292	10	BG212622	BG212622 RST32217
	33	214.8	39.9	245	10	BG208433	BG208433 RST27934
	34	213.6	39.7	849	10	BG436305	BG436305 602508742
	35	213.4	39.7	524	11	AF318372	AF318372 Homo sapi
c	36	207.6	38.6	507	9	AA972716	AA972716 op90e12.s
	37	207	38.5	256	10	BG216461	BG216461 RST36040
	38	207	38.5	547	10	N40147	N40147 yw73e12.r1
c	39	204.8	38.1	224	9	AA738394	AA738394 nx28c12.s
c	40	202.4	37.6	382	10	BF869799	BF869799 IL3-ET011
	41	200.6	37.3	499	10	BG354572	BG354572 PLAC6 Hum
c	42	193.8	36.0	197	9	AI968311	AI968311 wu01f07.x
c	43	192.8	35.8	226	10	BG220441	BG220441 RST40225
c	44	186.8	34.7	503	9	AI742551	AI742551 wg55c11.x
	45	185.8	34.5	756	10	BE903565	BE903565 601677770

ALIGNMENTS

RESULT 1
BG700165
LOCUS BG700165 569 bp mRNA linear EST 07-MAY-2001
DEFINITION 602679622F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812437 5', mRNA sequence.
ACCESSION BG700165
VERSION BG700165.1 GI:13969233
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 569)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10704 row: l column: 06
High quality sequence stop: 566.
FEATURES Location/Qualifiers
source 1. .569
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4812437"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and

normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 166 a 125 c 165 g 113 t
ORIGIN

Query Match 97.5%; Score 524.6; DB 10; Length 569;
Best Local Similarity 99.2%; Pred. No. 1e-111;
Matches 527; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      8 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     25 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 84

Qy     68 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACG 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     85 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACG 144

Qy    128 CTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGT 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    145 CTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGT 204

Qy    188 GGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGC 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    205 GGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGC 264

Qy    248 TCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAG 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    265 TCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATAG 324

Qy    308 CCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGA 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 CCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGA 384

Qy    368 GATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACA 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    385 GATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACA 444

Qy    428 GTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCAT 487
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    445 GTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCAT 504

Qy    488 TAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||| ||||| ||||||
Db    505 TAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAACAAAAAA 555
```

RESULT 2
BG773070
LOCUS BG773070 813 bp mRNA linear EST 15-MAY-2001
DEFINITION 602721370F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838176 5', mRNA sequence.
ACCESSION BG773070
VERSION BG773070.1 GI:14083723
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 813)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10771 row: l column: 17
High quality sequence stop: 778.
FEATURES Location/Qualifiers
source 1. .813
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838176"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 215 a 163 c 209 g 226 t
ORIGIN

Query Match 91.6%; Score 492.6; DB 10; Length 813;
Best Local Similarity 96.8%; Pred. No. 2.8e-104;
Matches 514; Conservative 0; Mismatches 14; Indels 3; Gaps 1;

```
Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 69

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
          ||||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||
Db     70 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTA 129

Qy    118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
          ||||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    130 GACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 189

Qy    178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 249

Qy    238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 309

Qy    298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AAGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 369

Qy    358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAG 429

Qy    418 AGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 477
          |||||||||||||||||||||||| | ||||||| ||||||||||||||||||||||||
Db    430 GGCAATCACAGTGTTAAAAGAAGACATGCTGAAATGTTGCAGGCTGCTCCTATGTTGGAA 489

Qy    478 ATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAA 528
          | || |||||| || ||||||||||||||||||||||||||||||||||||
Db    490 AATTCTTCATTGAAGTTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAA 540
```

RESULT 3
BI826605
LOCUS BI826605 527 bp mRNA linear EST 04-OCT-2001
DEFINITION 603077056F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5168892 5', mRNA sequence.
ACCESSION BI826605
VERSION BI826605.1 GI:15938155

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11419  row: p  column: 13
            High quality sequence stop: 519.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5168892"
                     /clone_lib="NIH_MGC_119"
                     /tissue_type="medulla"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source normal medulla from
                     anonymous male age 27.  Library  is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed  upon
                     cloning).  Average insert size 1.3 kb, insert size range
                     0.9-3 kb.  Library is normalized and enriched for
                     full-length clones and  was constructed by C. Gruber
                     (Invitrogen).  Research Genetics  tracking code 013. Note:
                     this is a NIH_MGC Library."
BASE COUNT      154 a    111 c    151 g    111 t
ORIGIN
```

```
  Query Match             91.4%;  Score 492;  DB 10;  Length 527;
  Best Local Similarity   98.3%;  Pred. No. 3.8e-104;
  Matches  519;  Conservative    0;  Mismatches    5;  Indels    4;  Gaps    2;

Qy      12 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 71
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCA 60

Qy      72 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGC 128
           ||||||||||||||||||||||||||||||||||   ||| |||||||||||||| |||
Db      61 TCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGC 120

Qy     129 TACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTG 188
           || ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTG 180

Qy     189 GAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCT 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCT 240

Qy     249 CAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGC 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGC 300

Qy     309 CAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAG 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAG 360

Qy     369 ATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAG 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAG 420

Qy     429 TGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATT 488
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATT 480

Qy     489 AAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 536
           |||||||||||||||||||||||| ||||||||||||||||||||
Db     481 AAAATTCTCCCAATAAAGCTTTACAG-CTTCTGCAAAGAAAAAAAAAA 527
```

```
RESULT   4
BG250953
LOCUS       BG250953                851 bp    mRNA    linear   EST 13-FEB-2001
DEFINITION  602363801F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4471969 5',
            mRNA sequence.
ACCESSION   BG250953
VERSION     BG250953.1  GI:12760769
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 851)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10292  row: b  column: 02
            High quality sequence stop: 531.
FEATURES             Location/Qualifiers
     source          1. .851
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4471969"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      224 a    205 c    239 g    183 t
ORIGIN
```

```
  Query Match             91.4%;  Score 492;  DB 10;  Length 851;
  Best Local Similarity   96.4%;  Pred. No. 3.9e-104;
  Matches  515;  Conservative    0;  Mismatches   15;  Indels    4;  Gaps    1;

Qy       8 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 70

Qy      68 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTA----TCGGCCTAGACCAA 123
           ||||||||||||||||||||||||||||||||||||||||    | | |    |||||
Db      71 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGCACCAA 130

Qy     124 GACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATG 183
           | ||||| ||| |||||||||| ||||||||||||||||||||||||||||||||||||
Db     131 GGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATG 190

Qy     184 AAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAG 243
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 AAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAG 250
```

```
Qy    244 CTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTC 303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    251 CTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTG 310

Qy    304 ATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGC 363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    311 ATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGC 370

Qy    364 AGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAAT 423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 AGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAAT 430

Qy    424 CACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGT 483
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    431 CACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGT 490

Qy    484 TCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
          |||||||||||||||||||||||||||||||||||||||| ||  ||||
Db    491 TCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAATAACCAAAA 544
```

RESULT 5
BG120838
LOCUS BG120838 572 bp mRNA linear EST 30-JAN-2001
DEFINITION 602351010F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4445882 5', mRNA sequence.
ACCESSION BG120838
VERSION BG120838.1 GI:12614347
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 572)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10224 row: c column: 03
High quality sequence start: 5
High quality sequence stop: 572.

```
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4445882"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      167 a    124 c    166 g    115 t
ORIGIN
```

Query Match 91.2%; Score 490.6; DB 10; Length 572;
Best Local Similarity 96.7%; Pred. No. 8.1e-104;
Matches 523; Conservative 0; Mismatches 14; Indels 4; Gaps 2;

```
Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     28 ACGCCAGGAGGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 87

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
          ||||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||
Db     88 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTA 147

Qy    118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
          ||||||| ||||| ||| |||||||||| ||||||||||||||||||||||||||||||
Db    148 GACCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCCGAGCAGTTCA 207

Qy    178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    208 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 267

Qy    238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    268 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 327

Qy    298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
          ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    328 AAGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 387

Qy    358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    388 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 447

Qy    418 AGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 477
          |||||||||||||||||||| |||||||||||||||||||| |||||||| ||||
Db    448 AGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTG-TCCTATGTCGGAA 506

Qy    478 ATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
          |||||||||||||||||||||||||||||||||||  |||||| ||| ||||||||||
Db    507 ATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACATGCTTCTGAAAAAAAAAAAAAAAA 566

Qy    538 A 538
          |
Db    567 A 567
```

RESULT 6
BI868671
LOCUS BI868671 509 bp mRNA linear EST 11-OCT-2001
DEFINITION 603392594F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5402663 5', mRNA sequence.
ACCESSION BI868671
VERSION BI868671.1 GI:16042344
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 509)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12026 row: l column: 24
High quality sequence stop: 509.

```
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5402663"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
```

/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 142 a 107 c 154 g 106 t
ORIGIN

Query Match 88.6%; Score 476.8; DB 10; Length 509;
Best Local Similarity 98.0%; Pred. No. 1.3e-100;
Matches 494; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

```
Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 ACGCTAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 65

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTA 117
          |||||||||||||||||||||||||||||||||||||||||||||    ||| ||||||
Db     66 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTA 125

Qy    118 GACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 177
          ||||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    126 GACCAAGGCGCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCA 185

Qy    178 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 237
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATC 245

Qy    238 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTG 297
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    246 CTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCTGAAGCCTG 305

Qy    298 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 AAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTG 365

Qy    358 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 417
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 ATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAA 425

Qy    418 AGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 477
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    426 AGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAA 485

Qy    478 ATTTGTTCATTAAAATTCTCCCAA 501
          |||||||||||||||||||||||
Db    486 ATTTGTTCATTAAAATTCTCCCAA 509
```

RESULT 7
BM172302
LOCUS BM172302 623 bp mRNA linear EST 04-DEC-2001
DEFINITION imageqc_4_2001/smu188bdff41.x1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4838176 5', mRNA sequence.
ACCESSION BM172302
VERSION BM172302.1 GI:17311865
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 623)
AUTHORS Kale,P.I., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and Prange,C.K.
TITLE The I.M.A.G.E. Consortium quality control effort: clone resequencing for verification
JOURNAL Unpublished (2001)
COMMENT Other_ESTs: BG773070
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been verified (found to hit its original self in the correct orientation), as part of the I.M.A.G.E. Consortium quality control effort. High quality sequence is defined as having 100 or more base pairs with a phred quality value of 20 or greater, where a sliding window of 4 base pairs with a phred quality value of 15 or greater marks the beginning and end of the sequence. For information on obtaining this clone, please contact info@image.llnl.gov.
Plate: LLAM10771 row: 1 column: 17
Seq primer: -21m13
High quality sequence stop: 623.
FEATURES Location/Qualifiers
source 1. .623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4838176"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 166 a 125 c 173 g 159 t
ORIGIN

Query Match 86.1%; Score 463.4; DB 10; Length 623;
Best Local Similarity 94.5%; Pred. No. 1.7e-97;
Matches 502; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

```
Qy      1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     10 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 69

Qy     61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTAT--CGGCCTAG 118
          ||||||||||||||||||||||||||||||||||||||||||||||||    | |
Db     70 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTA 129

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
              |  | ||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    130 TATCATGCTCTATGTACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 189

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    190 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 249

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    250 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 309

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    310 AGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 369

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    370 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAGG 429

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          |||||||||||||||||||| | ||||||| ||||||||||||||||||||||||||||
Db    430 GCAATCACAGTGTTAAAAGAAGACATGCTGAAATGTTGCAGGCTGCTCCTATGTTGGAAA 489

Qy    479 TTTGTTCA-TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAA 528
          || |||| || || |||||||||||||||||||||||||||||| ||||||
Db    490 ATTCTTCATTTGAAGTTCTCCCAATAAAGCTTTACAGCCTTCTGAAAAGAA 540
```

```
RESULT   8
AA447559/c
LOCUS       AA447559                489 bp    mRNA    linear   EST 04-JUN-1997
DEFINITION  zw81e11.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:782636
            3' similar to TR:G914905 G914905 GAGE-4 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AA447559
VERSION     AA447559.1  GI:2161229
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 489)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 427.
FEATURES             Location/Qualifiers
     source          1. .489
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:782636"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      101 a    139 c    105 g    144 t
ORIGIN
```

```
  Query Match             84.1%;  Score 452.4;  DB 9;  Length 489;
  Best Local Similarity   98.0%;  Pred. No. 6e-95;
  Matches  480;  Conservative    0;  Mismatches    6;  Indels    4;  Gaps    2;

Qy      43 CCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGAT 102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 CCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGAT 430

Qy     103 CGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTA 159
           |||||   ||| ||||||||||||| ||||| ||| |||||||||| ||||| ||||||
Db     429 CGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAAGTGATT-GGCCTA 371

Qy     160 TGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 219
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     370 TGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAG 311

Qy     220 CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 279
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     310 CAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAG 251

Qy     280 GTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 339
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     250 GTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTG 191

Qy     340 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 399
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     190 AGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAA 131

Qy     400 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAG 459
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     130 CGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAG  71

Qy     460 GCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTC 519
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70 GCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTC  11

Qy     520 TGCAAAGAAA 529
           ||||||||||
Db      10 TGCAAAGAAA 1
```

```
RESULT   9
AW510753/c
LOCUS       AW510753                457 bp    mRNA    linear   EST 03-MAR-2000
DEFINITION  hd39d05.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2911881 3' similar to SW:GGE4_HUMAN Q13068 GAGE-4 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AW510753
VERSION     AW510753.1  GI:7148831
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 457)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .457
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2911881"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT       87 a    133 c     95 g    142 t
ORIGIN
```

```
  Query Match             80.2%;  Score 431.4;  DB 9;  Length 457;
  Best Local Similarity   98.0%;  Pred. No. 4.5e-90;
  Matches  448;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy      77 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATC---GGCCTAGACCAAGACGCTACGT 133
           |||||||||||||||||||||||||||||||||    |||||||||||||| ||||| ||
Db     457 GTGAAATATGAGTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGT 398
```

```
Qy    134 AGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACC 193
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 ACAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACC 338

Qy    194 AGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGA 253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 AGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGA 278

Qy    254 GGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGA 313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 GGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGA 218

Qy    314 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGA 373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 ACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGA 158

Qy    374 CCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTA 433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 CCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTA 98

Qy    434 AAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAAT 493
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 AAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAAT 38

Qy    494 TCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAA 530
          |||||||||||||||||||||||||||||| ||||
Db     37 TCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAA 1
```

```
RESULT 10
AI381509/c
LOCUS       AI381509                447 bp    mRNA    linear   EST 28-MAR-1999
DEFINITION  te76b07.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGE:2092597 3' similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AI381509
VERSION     AI381509.1  GI:4194290
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 497   Std Error: 0.00
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2092597"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
```

```
BASE COUNT       84 a    130 c     95 g    138 t
ORIGIN
```

```
  Query Match             78.9%;  Score 424.6;  DB 9;  Length 447;
  Best Local Similarity   98.4%;  Pred. No. 1.7e-88;
  Matches  440;  Conservative    0;  Mismatches    4;  Indels    3;  Gaps    1;

Qy     88 GTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTG 144
          |||||||||||||||||||    ||| |||||||||||| ||||| ||| |||||||||
Db    447 GTTGGCGAGGAAGATCGACCTATTATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTG 388

Qy    145 AAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTG 204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 AAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTG 328

Qy    205 AAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATG 264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 AAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATG 268

Qy    265 AGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACC 324
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 AGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACC 208

Qy    325 CACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATC 384
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 CACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATC 148

Qy    385 CAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACAC 444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 CAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACAC 88

Qy    445 GTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAA 504
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 GTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAA 28

Qy    505 AGCTTTACAGCCTTCTGCAAAGAAAAA 531
          |||||||||||||||||||||||||||
Db     27 AGCTTTACAGCCTTCTGCAAAGAAAAA 1
```

```
RESULT 11
AA738037/c
LOCUS       AA738037                464 bp    mRNA    linear   EST 22-JAN-1998
DEFINITION  nx15e11.s1 NCI_CGAP_GC3 Homo sapiens cDNA clone IMAGE:1256204 3'
            similar to TR:Q13070 Q13070 GAGE-6 PROTEIN. ;, mRNA sequence.
ACCESSION   AA738037
VERSION     AA738037.1  GI:2768794
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 511   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 435.
FEATURES             Location/Qualifiers
     source          1. .464
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1256204"
                     /clone_lib="NCI_CGAP_GC3"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from 3 pooled
                     germ cell tumors, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  Library is not normalized. Library was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       86 a    132 c     93 g    153 t
ORIGIN


  Query Match             78.9%;  Score 424.4;  DB 9;  Length 464;
  Best Local Similarity   96.9%;  Pred. No. 1.9e-88;
  Matches  444;  Conservative    0;  Mismatches   11;  Indels    3;  Gaps    1;

Qy      84 ATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGACGCTACGTAGAGCCT 140
           |||||||||||||||||||||||   ||| ||||||| |||||||||| ||| |||||
Db     464 ATGAGTTGGCGAGGAAGATCGACCTGTTATTGGCCTAGTCCAAGACGCTATGTACAGCCT 405

Qy     141 CCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 200
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     404 CCTGAAATGATTGGGCCTATGCAGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACA 345

Qy     201 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 CCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAG 285

Qy     261 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGT 320
           |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     284 GATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGATAGCCAGGAACAGGGT 225

Qy     321 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 CACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCA 165

Qy     381 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     164 AATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAG 105

Qy     441 ACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     104 GCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCA 45

Qy     501 ATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||   |||||||||||||||| ||||||||||||
Db      44 ATAAGAGTTTACAGCCTTCTGCAAAAAAAAAAAAAAA 7


RESULT  12
BG700088
LOCUS       BG700088                581 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602679431F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812462 5',
            mRNA sequence.
ACCESSION   BG700088
VERSION     BG700088.1  GI:13969078
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10704  row: m  column: 07
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4812462"
                     /clone_lib="NIH_MGC_95"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT      164 a    142 c    166 g    109 t
ORIGIN


  Query Match             78.8%;  Score 424;  DB 10;  Length 581;
  Best Local Similarity   92.8%;  Pred. No. 2.4e-88;
  Matches  501;  Conservative    0;  Mismatches   30;  Indels    9;  Gaps    5;

Qy       8 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 67
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25 GGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGA 84

Qy      68 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACG 127
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      85 TTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGACCAAGACG 144

Qy     128 CTACGTAGAGCCTCCTGA-AATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     145 CTACGTAGAGCCTCCTGACAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 204

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     205 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 264

Qy     247 CTCAGGAGGGAG-AGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCAT 305
           |||||||||||| ||| ||||||||||||||||||||||||||||||||||||| ||
Db     265 CTCAGGAGGGAGCAGGCTGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTGAT 324

Qy     306 AGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAG 365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     325 AGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAG 384

Qy     366 GAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGA-AGGTGAAAAGCAATC 424
           |||||||||||| |||||||||||||| |||||||||||||||| ||||||||||||||
Db     385 GAGATGGACCCGGCAAATCCAGAGGACGTGAAAACGCCTGAAGACAGGTGAAAAGCAATC 444

Qy     425 ACA---GTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 481
           |||   ||   |||  | |   |||||| |||||||||||||||||||||||||||| |
Db     445 ACAAGTGTTACAAAGAAGGCACCGTTGACATGATGCAGGCTGCTCCTATGTTGGAAACTC 504

Qy     482 GTTCA---TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
             | |   |  | |||||||| ||||||| |||||||||||||||  ||| |  | |||
Db     505 TGTAACATTACACATTCTCCCACTAAAGCTATACAGCCTTCTGCAACAAAACACCACAAA 564


RESULT  13
```

```
AW016546/c
LOCUS       AW016546                418 bp    mRNA    linear   EST 10-SEP-1999
DEFINITION  UI-H-BI0p-abg-g-06-0-UI.s1 NCI_CGAP_Sub2 Homo sapiens cDNA clone
            IMAGE:2711986 3', mRNA sequence.
ACCESSION   AW016546
VERSION     AW016546.1  GI:5865303
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 418)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .418
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2711986"
                     /clone_lib="NCI_CGAP_Sub2"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The
                     NCI_CGAP_Sub2 library is a subtracted library derived from
                     BI. BI constitutes a mixture of 21 normalized or
                     subtracted NCI_CGAP libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22
                     , NCI_CGAP_Pr28, NCI_CGAP_Co10, NCI_CGAP_Co16,
                     NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3,
                     NCI_CGAP_Kid11, NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8,
                     NCI_CGAP_CLL1, NCI_CGAP_Lei2, NCI_CGAP_Brn23, NCI_CGAP_Lu5
                     , NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6
                     , NCI_CGAP_Brn25.  These 21 libraries were pooled and a
                     single-stranded DNA preparation of the resulting mixture
                     was used as a tracer in a subtractive hybridization with a
                     driver whose composition is detailed below: NCI_CGAP_Kid3
                     pool 1 LLAM 3334-3337, 3682-3683, 3798-3803   (IMAGE
                     CloneIDs 1322376-1323911, 1456008-1456775, 1500552-1502855
                     ) NCI_CGAP_Kid5 pool 1 LLAM 3338-3342, 3722-3725,
                     3776-3778   (IMAGE CloneIDs 1323912-1325831,
                     1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LLAM
                     3575-3582, 3851-3854   (IMAGE CloneIDs 1414920-1417991,
                     1520904-1522439)  NCI_CGAP_GC4 pool 1 LLAM 3164-3167,
                     3716-3720, 3733-3735   (IMAGE CloneIDs 1257096-1258631,
                     1469064-1470983, 1475592-1476743) NCI_CGAP_Pr22 pool 1
                     LLAM 2457-2459, 2758-2759, 3062-3068   (IMAGE CloneIDs
                     985608-986759, 1101192-1101959, 1217928-1220615)
                     NCI_CGAP_Co10 pool 1 LLAM 2644-2653, 2871-2872   (IMAGE
                     CloneIDs 1057416-1061255, 1144584-1145351)  The resulting
                     subtracted library contained 4 million recombinants.
                     Subtraction was performed as previously described [Bonaldo
                     ,  Lennon & Soares (1996): Normalization and Subtraction:
                     Two Approaches  To Facilitate Gene Discovery. Genome
                     Research 6, 791-806.
                     TAG_LIB=NCI_CGAP_GC4
                     TAG_TISSUE=germ cell
                     TAG_SEQ=AAATC"
BASE COUNT       74 a    119 c     87 g    138 t
ORIGIN
```

```
  Query Match           76.9%;  Score 413.8;  DB 9;  Length 418;
  Best Local Similarity   99.5%;  Pred. No. 5.5e-86;
  Matches  415;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    122 AAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    418 AAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGA 359

Qy    182 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358 TGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGC 299

Qy    242 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 AGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGC 239

Qy    302 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 361
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238 TCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGG 179

Qy    362 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178 GCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCA 119

Qy    422 ATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 481
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    118 ATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTT 59

Qy    482 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     58 GTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAAAAAAAAAAAAA 2
```

```
RESULT  14
BG120336
LOCUS       BG120336                480 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602353732F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4451958 5',
            mRNA sequence.
ACCESSION   BG120336
VERSION     BG120336.1  GI:12613845
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 480)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10239  row: p  column: 07
            High quality sequence stop: 480.
FEATURES             Location/Qualifiers
     source          1. .480
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4451958"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
```

```
BASE COUNT      170 a     93 c    133 g     84 t
ORIGIN


  Query Match             76.8%;  Score 413.4;  DB 10;  Length 480;
  Best Local Similarity   97.4%;  Pred. No. 6.8e-86;
  Matches  420;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    108 TATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCC 167
          ||| |||||||||||||| ||||| ||| |||||||||| ||||||||||||||||||||
Db     15 TATTGGCCTAGACCAAGGCGCTATGTACAGCCTCCTGAAGTGATTGGGCCTATGCGGCCC 74

Qy    168 GAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAA 227
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 GAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAA 134

Qy    228 CGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGG 287
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 CGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGG 194

Qy    288 CCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAA 347
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    195 CCGAAGCCTGAAGCTGATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAA 254

Qy    348 GATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAA 407
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 GATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAA 314

Qy    408 GAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCC 467
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    315 GAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCC 374

Qy    468 TATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGA 527
          |||||||||||||||||||||||||||||||||||||||||||||||||||   || | |
Db    375 TATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTTGCAAAA 434

Qy    528 AAAAAAAAAAA 538
          ||||||||||| |
Db    435 AAAAAAAAAAA 445


RESULT  15
AA868226/c
LOCUS       AA868226                426 bp    mRNA    linear   EST 16-MAR-1998
DEFINITION  ak48h07.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1409245
            3' similar to SW:GGE2_HUMAN Q13066 GAGE-2 PROTEIN. [1] ;, mRNA
            sequence.
ACCESSION   AA868226
VERSION     AA868226.1  GI:2963671
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Possible reversed clone: similarity on wrong strand
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 363.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1409245"
                     /clone_lib="Soares_testis_NHT"
                     /sex="male"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. 
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT       79 a    119 c     95 g    131 t      2 others
ORIGIN


  Query Match             76.7%;  Score 412.8;  DB 9;  Length 426;
  Best Local Similarity   97.9%;  Pred. No. 9.4e-86;
  Matches  417;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy    102 TCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATG 161
          |||||||||||||||||||||||||||||||||||||||||||||||||||     |||||||
Db    426 TCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGNANTTGCCTATG 367

Qy    162 CGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCA 221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    366 CGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCA 307

Qy    222 ACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGT 281
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    306 ACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGT 247

Qy    282 CAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAG 341
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAG 187

Qy    342 TGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACG 401
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 TGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACG 127

Qy    402 CCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGC 461
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    126 CCTGAAGAAGGTGAAAAGCAATCACAGTGTTAAAAGAAGGCACGTTGAAATGATGCAGGC 67

Qy    462 TGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTG 521
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     66 TGCTCCTATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAACTTTACAGCCTTCTT 7

Qy    522 CAAAGA 527
          |||| |
Db      6 CAAAAA 1


Search completed: July  1, 2002, 06:39:18
Job time: 1653 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2002, 06:11:45 ; Search time 201.18 Seconds
(without alignments)
4591.411 Million cell updates/sec

Title: US-09-782-745-14
Perfect score: 538
Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAAAAAAA 538

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	538	100.0	538	20	AAX90519	GAGE-2 tumour reje
2	531.6	98.8	1245	22	AAS60826	Human cancer agent
3	512.8	95.3	540	20	AAX90521	GAGE-4 tumour reje
4	510.2	94.8	540	20	AAX90523	GAGE-6 tumour reje
5	504.8	93.8	532	20	AAX90522	GAGE-5 tumour reje
6	500.6	93.0	539	19	AAV18721	cDNA encoding GAGE
7	496.8	92.3	532	19	AAV18720	cDNA encoding GAGE
8	493	91.6	535	19	AAV18717	cDNA encoding GAGE
9	490.6	91.2	541	19	AAV18719	cDNA encoding GAGE
10	429.2	79.8	560	20	AAX90520	GAGE-3 tumour reje
11	421.2	78.3	560	19	AAV18718	cDNA encoding GAGE
12	399.4	74.2	1024	21	AAZ97217	Human prostate can
c 13	397.4	73.9	1024	21	AAZ97216	Human prostate can
14	374.4	69.6	646	17	AAT31861	GAGE tumour reject
15	374.4	69.6	646	20	AAX90518	GAGE tumour reject
16	368	68.4	646	19	AAV05540	cDNA encoding GAGE
17	366.4	68.1	648	16	AAQ81405	GAGE tumour reject
18	357.8	66.5	430	21	AAC02129	Human secreted pro
c 19	301.4	56.0	530	22	AAS60104	Human cancer agent
c 20	299.2	55.6	365	22	AAS60496	Human cancer agent
21	203.6	37.8	650	21	AAA59322	cDNA encoding a ca
22	194	36.1	532	22	AAI60530	Human polynucleoti
23	189.2	35.2	659	22	AAI58744	Human polynucleoti
24	181.8	33.8	673	22	AAK51918	Human polynucleoti
25	173.8	32.3	611	22	AAD14983	Human NOV4 DNA. H
26	172.4	32.0	503	23	AAS69484	DNA encoding novel
27	165.6	30.8	656	21	AAZ97402	Human prostate can
28	156	29.0	580	22	AAF59637	Human cell cycle a
29	139.8	26.0	661	23	AAS69486	DNA encoding novel
30	131.4	24.4	11462	19	AAV05532	Nucleic acid seque
31	130.4	24.2	475	22	AAD14981	Human NOV2 DNA. H
c 32	127.8	23.8	665	22	AAK52902	Human polynucleoti
33	120.8	22.5	1051	22	AAD14982	Human NOV3 DNA. H
34	111.4	20.7	762	22	AAH64751	Human secreted pro
35	108.4	20.1	781	22	AAF68861	Human lung tumour
36	108	20.1	399	22	AAF68151	Human lung tumour
37	105.4	19.6	750	23	AAS91235	DNA encoding novel
38	105.2	19.6	571	21	AAZ33316	Human secreted pro
39	101.6	18.9	515	22	AAS63900	Human prostate cDN
40	101.6	18.9	515	22	AAH93807	Human prostate-spe
41	101.6	18.9	515	22	AAH85121	Human prostate-spe
42	101.6	18.9	515	22	AAH02872	Prostate tumour an
43	100.8	18.7	214	22	AAS24637	Human ovarian PCR-
44	100.8	18.7	320	22	AAH83260	Human ovarian tumo
c 45	93.4	17.4	472	21	AAZ95012	CSG Pro118 EST use

ALIGNMENTS

RESULT 1
AAX90519
ID AAX90519 standard; cDNA; 538 BP.
XX
AC AAX90519;
XX
DT 30-SEP-1999 (first entry)
XX
DE GAGE-2 tumour rejection antigen clone nucleotide sequence.
XX
KW Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW detection; therapy; pathological condition; cancer; CTL;
KW cytolytic T lymphocyte; GAGE; ss.
XX
OS Homo sapiens.
XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX
DR WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of

PT pathological conditions, e.g. cancer
XX
PS Example 13; Fig 4; 62pp; English.
XX
CC The present invention describes peptides which bind to human leukocyte
CC antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC tumour rejection antigens. They can be used for detecting cytolytic T
CC lymphocytes (CTLs) in pathological conditions such as cancer and in
CC HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC be used for stimulating CTLs in vivo. The present sequence represents
CC a GAGE tumour rejection antigen clone, from an example from the present
CC invention.
XX
SQ Sequence 538 BP; 160 A; 116 C; 155 G; 107 T; 0 other;

Query Match 100.0%; Score 538; DB 20; Length 538;
Best Local Similarity 100.0%; Pred. No. 3e-146;
Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 acgccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttttcctct 60

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 actgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctatcggcctagac 120

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 caagacgctacgtagagcctcctgaaatgattgggcctatgcggcccgagcagttcagtg 180

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 atgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctg 240

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 cagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaag 300

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ctcatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatg 360

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagc 420

Qy     421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aatcacagtgttaaaagaagacacgttgaaatgatgcaggctgctcctatgttggaaatt 480

Qy     481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 tgttcattaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 538
```

RESULT 2
AAS60826
ID AAS60826 standard; cDNA; 1245 BP.
XX
AC AAS60826;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human cancer agent-sensitive marker #328.
XX
KW Human; cancer cell marker; TAXOL; cytostatic; tumour; carcinoma;
KW squamous cell carcinoma; sarcoma; fibrosarcoma; leukaemia;
KW lymphocytic leukaemia; lymphoma; plasmocytoma; reticulum cell sarcoma;
KW Hodgkin's disease; glioma; ss.
XX
OS Homo sapiens.

XX
PN WO200179556-A2.
XX
PD 25-OCT-2001.
XX
PF 13-APR-2001; 2001WO-US12132.
XX
PR 14-APR-2000; 2000US-197538P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Brown JL, Bolt A, Van Huffel C;
XX
DR WPI; 2001-602933/68.
XX
PT Novel nucleic acid, used as a marker to determine the effectiveness of
PT using TAXOL to treat cancer cell growth in individuals -
XX
PS Claim 1; Page 340; 527pp; English.
XX
CC The invention relates to 1046 novel nucleic acids which are used as
CC markers for determining the sensitivity of a cancer cell to the
CC anticancer agent TAXOL. Cancer cells can be treated with TAXOL when
CC they are shown to express one of the 242 sensitivity markers or the
CC cells are shown not to express one of the 804 resistance markers.
CC The methods can be used to determine the effectiveness of TAXOL
CC in the treatment of cancer cell growth in an individual. The markers
CC can be used as targets in developing anti-cancer agents such as
CC chemotherapeutic compounds. The markers can also be used as targets in
CC developing treatments for cancer, particularly those cancers which
CC display resistance to agents and exhibit expression of the markers. The
CC anticancer agents developed by the novel method can be used to treat
CC cancer. Probes based on the markers can be used to detect transcripts or
CC genomic sequences corresponding to the markers, in the identification of
CC cells or tissues which mis-express the protein. Cancers which may
CC be targeted include carcinoma (e.g. squamous cell carcinoma),
CC sarcoma (e.g. fibrosarcoma) leukaemia (e.g. lymphocytic leukaemia),
CC lymphoma, plasmocytoma, reticulum cell sarcoma, Hodgkin's disease and
CC tumours (e.g. glioma). The present sequence is one of the 1046
CC novel cancer cell markers.
XX
SQ Sequence 1245 BP; 382 A; 308 C; 390 G; 153 T; 12 other;

Query Match 98.8%; Score 531.6; DB 22; Length 1245;
Best Local Similarity 99.3%; Pred. No. 3.2e-144;
Matches 534; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     179 acgccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctct 238

Qy      61 ACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db     239 actgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctattggcctagac 298

Qy     121 CAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     299 caagacgctacgtagagcctcctgaaatgattgggcctatgcggcccgagcagttcagtg 358

Qy     181 ATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     359 atgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctg 418

Qy     241 CAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     419 cagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaag 478

Qy     301 CTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     479 ctcatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatg 538

Qy     361 GGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGC 420
```

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    539 ggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagc 598

Qy    421 AATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATT 480
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    599 aatcacagtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatt 658

Qy    481 TGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||| |||||||||| |
Db    659 tgttcattaaaattctcccaataaagctttacagccttctgcaaaaaaaaaaaaaaga 716


RESULT   3
AAX90521
ID   AAX90521 standard; cDNA; 540 BP.
XX
AC   AAX90521;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE-4 tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937665-A1.
XX
PD   29-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00775.
XX
PR   23-JAN-1998;   98US-0012818.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B,  Van Der Bruggen P;
XX
DR   WPI; 1999-469111/39.
XX
PT   New isolated peptides which bind to HLA-A29 molecules, which are
PT   tumour rejection antigens used for detection and therapy of
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
SQ   Sequence 540 BP; 159 A; 114 C; 156 G; 111 T; 0 other;


  Query Match             95.3%;  Score 512.8;  DB 20;  Length 540;
  Best Local Similarity   98.1%;  Pred. No. 6.2e-139;
  Matches  530;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cgccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctcta 60

Qy     62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
          |||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db     61 ctgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctag 120

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
          |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db    121 accaaggcgctatgtacagcctcctgaaatgattgggcctatgcggcccgagcagttcag 180

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 tgatgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcc 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tgcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctga 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 agctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctga 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tgggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaa 420

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAA 478
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    421 gcaatcacagtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaa 480

Qy    479 TTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    481 tttgttcattaaaattctcccaataaagctttacagccttctgcaaaaaaaaaaaaaaaa 540


RESULT   4
AAX90523
ID   AAX90523 standard; cDNA; 540 BP.
XX
AC   AAX90523;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE-6 tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937665-A1.
XX
PD   29-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00775.
XX
PR   23-JAN-1998;   98US-0012818.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B,  Van Der Bruggen P;
XX
DR   WPI; 1999-469111/39.
XX
PT   New isolated peptides which bind to HLA-A29 molecules, which are
PT   tumour rejection antigens used for detection and therapy of
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
```

```
SQ   Sequence 540 BP; 159 A; 113 C; 157 G; 111 T; 0 other;

  Query Match             94.8%;  Score 510.2;  DB 20;  Length 540;
  Best Local Similarity   98.0%;  Pred. No. 3.6e-138;
  Matches  528;  Conservative    0;  Mismatches    8;  Indels    3;  Gaps    1;

Qy       3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctctac 60

Qy      63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
          |||||||||||||||||||||||||||||||||||||||||||||   ||| ||||||||
Db      61 tgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctaga 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
          ||||| ||||| ||| ||||||||||| ||||||||||||||||||||||||||||||||
Db     121 ccaaggcgctatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagt 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 gatgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcct 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaa 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgat 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gggcaggaggtggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaag 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 caatcacagtgttaaaagaagacacgttgaaatgatgcaggctgctcctatgttggaaat 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 ttgttcattaaaattctcccaataaagctttacagccttctgcaaaaaaaaaaaaaaaa 539


RESULT   5
AAX90522
ID   AAX90522 standard; cDNA; 532 BP.
XX
AC   AAX90522;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE-5 tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937665-A1.
XX
PD   29-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00775.
XX
PR   23-JAN-1998;   98US-0012818.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX
DR   WPI; 1999-469111/39.
XX
PT   New isolated peptides which bind to HLA-A29 molecules, which are
PT   tumour rejection antigens used for detection and therapy of
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
SQ   Sequence 532 BP; 156 A; 111 C; 154 G; 111 T; 0 other;

  Query Match             93.8%;  Score 504.8;  DB 20;  Length 532;
  Best Local Similarity   98.1%;  Pred. No. 1.3e-136;
  Matches  522;  Conservative    0;  Mismatches    7;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctctactgagatt 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           ||||||||||||||||||||||||||||||||||||   ||| |||||||||||| |
Db      61 catctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctagaccaaggc 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           |||| ||| |||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     121 gctatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagtgatgaag 180

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 tggaaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctgcagctg 240

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     241 ctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaagctgata 300

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatgggcagg 360

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 agatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagcaatcac 420

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db     421 agtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatttgttca 480

Qy     487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ttaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 532


RESULT   6
AAV18721
ID   AAV18721 standard; cDNA; 539 BP.
XX
AC   AAV18721;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-6 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
```

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             82..436
FT                   /*tag= a
FT                   /transl_except= (pos:127..129, aa:Arg)
FT                   /transl_except= (pos:196..198, aa:Ala)
FT                   /transl_except= (pos:199..201, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47603.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-6 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 539 BP; 157 A; 114 C; 157 G; 111 T; 0 other;


  Query Match             93.0%;  Score 500.6;  DB 19;  Length 539;
  Best Local Similarity   96.8%;  Pred. No. 2.2e-135;
  Matches  522;  Conservative    0;  Mismatches   14;  Indels    3;  Gaps    1;

Qy       3 GCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTAC 62
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gccagggagctgtgaggcagtgctgtgtggttcctgccgtccggactcttttcctctac 60

Qy      63 TGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGA 119
           ||||||||||||||||||||||||||||||||||||||||||||    ||| ||||||||
Db      61 tgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctaga 120

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||| | ||| ||| |||||||||||| ||||||||||||||||||||||||||||||
Db     121 ccaaggcactatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagt 180

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||   | ||||||||||||||||||||||||||||||||||||||
Db     181 gatgaagtggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcct 240

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 gcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaa 300

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgat 360

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gggcaggaggtggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaag 420

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     421 caatcacagtgttaaaagaagacacgctgaaatgatgcaggctgctcctatgttggaaat 480

Qy     480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     481 ttgttcattaaaattctcccaataaagctttacagccttctgcaaaaaaaaaaaaaaaa 539


RESULT   7
AAV18720
ID   AAV18720 standard; cDNA; 532 BP.
XX
AC   AAV18720;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-5 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             75..429
FT                   /*tag= a
FT                   /transl_except= (pos:189..191, aa:Ala)
FT                   /transl_except= (pos:192..194, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47602.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-5 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
```

```
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 532 BP; 154 A; 110 C; 156 G; 112 T; 0 other;
```

```
  Query Match             92.3%;  Score 496.8;  DB 19;  Length 532;
  Best Local Similarity   97.2%;  Pred. No. 2.7e-134;
  Matches  517;  Conservative    0;  Mismatches   12;  Indels    3;  Gaps    1;

Qy      10 AGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTACTGAGATT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 agctgtgaggcagtgctgtgtggttcctgccgtccggactcttttttcctctactgagatt 60

Qy      70 CATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAGACCAAGAC 126
           |||||||||||||||||||||||||||||||||||||   ||| |||||||||||| |
Db      61 catctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctagaccaaggc 120

Qy     127 GCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAG 186
           | || ||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     121 ggtatgtacagcctcctgaagtgattgggcctatgcggcccgagcagttcagtgatgaag 180

Qy     187 TGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTG 246
           |||||||||   | ||||||||||||||||||||||||||||||||||||||||||||
Db     181 tggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcctgcagctg 240

Qy     247 CTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATA 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db     241 ctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaagctgata 300

Qy     307 GCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatgggcagg 360

Qy     367 AGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCAC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 agatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagcaatcac 420

Qy     427 AGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCA 486
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     421 agtgttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatttgttca 480

Qy     487 TTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ttaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaa 532
```

```
RESULT   8
AAV18717
ID   AAV18717 standard; cDNA; 535 BP.
XX
AC   AAV18717;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-2 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             81..431
FT                   /*tag=  a
FT                   /transl_except= (pos:192..194, aa:Ala)
FT                   /transl_except= (pos:195..197, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47599.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-2 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 535 BP; 158 A; 112 C; 156 G; 109 T; 0 other;
```

```
  Query Match             91.6%;  Score 493;  DB 19;  Length 535;
  Best Local Similarity   97.2%;  Pred. No. 3.5e-133;
  Matches  524;  Conservative    0;  Mismatches   10;  Indels    5;  Gaps    2;

Qy       1 ACGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTC- 59
           |||||||||||||||||||||||||||||    | ||  ||||||||||||||||||| 
Db       1 acgccagggagctgtgaggcagtgctgtg----ttcttgcgtccggactctttttcctct 56

Qy      60 TACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCTATCGGCCTAGA 119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57 tactgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctatcggcctaga 116

Qy     120 CCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGT 179
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     117 ccaagacgctacgtagagcctcctgaaatgattgggcctatgcggcccgagcagttcagt 176

Qy     180 GATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCT 239
           ||||||||||||||||   | ||||||||||||||||||||||||||||||||||||||
Db     177 gatgaagtggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcct 236

Qy     240 GCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAA 299
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     237 gcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaa 296

Qy     300 GCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGAT 359
           ||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     297 gctgagagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgat 356

Qy     360 GGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAG 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     357 gggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaag 416

Qy     420 CAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAAT 479
```

```
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    417 caatcacagtgttaaaagaagacatgttgaaatgatgcaggctgctcctatgttggaaat 476

Qy    480 TTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 ttgttcattaaaattctcccaataaagctttacagccttctgcaaagaaaaaaaaaaaaa 535


RESULT  9
AAV18719
ID   AAV18719 standard; cDNA; 541 BP.
XX
AC   AAV18719;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-4 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             83..435
FT                   /*tag= a
FT                   /transl_except= (pos:197..199, aa:Ala)
FT                   /transl_except= (pos:200..202, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47601.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-4 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 541 BP; 158 A; 117 C; 154 G; 112 T; 0 other;


  Query Match             91.2%;  Score 490.6;  DB 19;  Length 541;
  Best Local Similarity   96.7%;  Pred. No. 1.8e-132;

  Matches  523;  Conservative    0;  Mismatches   14;  Indels    4;  Gaps    2;

Qy      2 CGCCAGGGAGCTGTGAGGCAGTGCTGTGTGGTTCCTGCCGTCCGGACTCTTTTTCCTCTA 61
          ||||| |||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db      1 cgccacggagctgtgaggcagtgctgtgtggttcctgccgtccccactctttttcctcta 60

Qy     62 CTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACC---TATCGGCCTAG 118
          ||||||||||||||||||||||||||||||||||||||||||||||   ||| |||||||
Db     61 ctgagattcatctgtgtgaaatatgagttggcgaggaagatcgacctattattggcctag 120

Qy    119 ACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAG 178
          |||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    121 accaaggcgctatgtacagcctcctgaaatgattgggcctatgcggcccgagcagttcag 180

Qy    179 TGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCC 238
          |||||||||||||||||   | ||||||||||||||||||||||||||||||||||||
Db    181 tgatgaagtggaaccagagcctcctgaagaaggggaaccagcaactcaacgtcaggatcc 240

Qy    239 TGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGA 298
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tgcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggccgaagcctga 300

Qy    299 AGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGA 358
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 agctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctga 360

Qy    359 TGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAA 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 tgggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaa 420

Qy    419 GCAATCACAGTGTTAAAAGAAGACACGTTG-AAATGATGCAGGCTGCTCCTATGTTGGAA 477
          ||||||||||||||||||| ||||||| ||||||||||||||||||||||||||||||
Db    421 gcaatcacagtgttaaaagaaggcacgttgaaaatgatgcaggctgctcctatgttggaa 480

Qy    478 ATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAA 537
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    481 atttgttcattaaaattctcccaataaagctttacagccttctgcaaaaaaaaaaaaaaa 540

Qy    538 A 538
          |
Db    541 a 541


RESULT  10
AAX90520
ID   AAX90520 standard; cDNA; 560 BP.
XX
AC   AAX90520;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE-3 tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; ss.
XX
OS   Homo sapiens.
XX
PN   WO9937665-A1.
XX
PD   29-JUL-1999.
XX
PF   12-JAN-1999;   99WO-US00775.
XX
PR   23-JAN-1998;   98US-0012818.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B,  Van Der Bruggen P;
XX
DR   WPI; 1999-469111/39.
XX
```

```
PT   New isolated peptides which bind to HLA-A29 molecules, which are
PT   tumour rejection antigens used for detection and therapy of
PT   pathological conditions, e.g. cancer
XX
PS   Example 13; Fig 4; 62pp; English.
XX
CC   The present invention describes peptides which bind to human leukocyte
CC   antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC   tumour rejection antigens. They can be used for detecting cytolytic T
CC   lymphocytes (CTLs) in pathological conditions such as cancer and in
CC   HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC   be used for stimulating CTLs in vivo. The present sequence represents
CC   a GAGE tumour rejection antigen clone, from an example from the present
CC   invention.
XX
SQ   Sequence 560 BP; 177 A; 119 C; 150 G; 114 T; 0 other;
```

```
  Query Match             79.8%;  Score 429.2;  DB 20;  Length 560;
  Best Local Similarity   92.2%;  Pred. No. 1.2e-114;
  Matches  452;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      49 TCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCT 108
           ||||  ||  | ||| | |||   |   |   ||| | ||| |  ||  |    |    |
Db      71 tcttcctcgccaactcatatttcacacagatgaatctcagtagaggaaaatcgacctatt 130

Qy     109 ATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCG 168
           || ||||||||||||| ||||| ||| |||||||||| ||||||||||||||||||||||
Db     131 attggcctagaccaaggcgctatgtacagcctcctgaagtgattgggcctatgcggcccg 190

Qy     169 AGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAAC 228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     191 agcagttcagtgatgaagtggaaccagcaacacctgaagaaggggaaccagcaactcaac 250

Qy     229 GTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGC 288
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 gtcaggatcctgcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggc 310

Qy     289 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAG 348
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     311 cgaagcctgaagctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaag 370

Qy     349 ATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAG 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     371 atggtcctgatgggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaag 430

Qy     409 AAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCT 468
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     431 aaggtgaaaagcaatcacagtgttaaaagaaggcacgttgaaatgatgcaggctgctcct 490

Qy     469 ATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAA 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     491 atgttggaaatttgttcattaaaattctcccaataaagctttacagccttctgcaaagaa 550

Qy     529 AAAAAAAAAA 538
           ||||||||||
Db     551 aaaaaaaaaa 560
```

```
RESULT  11
AAV18718
ID   AAV18718 standard; cDNA; 560 BP.
XX
AC   AAV18718;
XX
DT   30-JUL-1998  (first entry)
XX
DE   cDNA encoding GAGE-3 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             100..457
FT                   /*tag=  a
FT                   /transl_except= (pos:217..219, aa:Ala)
FT                   /transl_except= (pos:220..222, aa:Thr)
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   P-PSDB; AAW47600.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 4; 60pp; English.
XX
CC   The present sequence encodes a GAGE-3 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein is
CC   testis. Several GAGE TRAPs have been identified (see AAV18717-21). The
CC   major difference between these proteins and GAGE-1 (AAV05540) is the
CC   absence of a stretch of 143 bases located at position 379 to 521 of the
CC   GAGE-1 TRAP sequence. The rest of the sequences show mismatches at
CC   various position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence 560 BP; 175 A; 119 C; 152 G; 114 T; 0 other;
```

```
  Query Match             78.3%;  Score 421.2;  DB 19;  Length 560;
  Best Local Similarity   91.2%;  Pred. No. 2.6e-112;
  Matches  447;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      49 TCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGAGGAAGATCGACCT 108
           || |  ||  | ||| | |||   |   |   ||| | ||| |  ||  |    |    |
Db      71 tcgtcctcgccaactcatatttcacacagatgaatctcagtagaggaaaatcgacctatt 130

Qy     109 ATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCG 168
           || ||||||||||||| ||||| ||| |||||||||| ||||||||||||||||||||||
Db     131 attggcctagaccaaggcgctatgtacagcctcctgaagtgattgggcctatgcggcccg 190

Qy     169 AGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAAC 228
           |||||||||||||||||||||||||||   | ||||||||||||||||||||||||||||||
Db     191 agcagttcagtgatgaagtggaaccagagcctcctgaagaaggggaaccagcaactcaac 250

Qy     229 GTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGC 288
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     251 gtcaggatcctgcagctgctcaggagggagaggatgagggagcatctgcaggtcaagggc 310

Qy     289 CGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAG 348
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     311 cgaagcctgaagctgatagccaggaacagggtcacccacagactgggtgtgagtgtgaag 370
```

```
Qy    349 ATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAG 408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    371 atggtcctgatgggcaggagatggacccgccaaatccagaggaggtgaaaacgcctgaag 430

Qy    409 AAGGTGAAAAGCAATCACAGTGTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCT 468
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    431 aaggtgaaaagcaatcacagtgttaaaagaaggcacgttgaaatgatgcaggctgctcct 490

Qy    469 ATGTTGGAAATTTGTTCATTAAAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAA 528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    491 atgttggaaatttgttcattaaaattctcccaataaagctttacagccttctgcaaagaa 550

Qy    529 AAAAAAAAAA 538
          ||||||||||
Db    551 aaaaaaaaaa 560
```

RESULT 12
AAZ97217
ID AAZ97217 standard; cDNA; 1024 BP.
XX
AC AAZ97217;
XX
DT 18-APR-2000 (first entry)
XX
DE Human prostate cancer differentially expressed gene #78.
XX
KW Prostate cancer specific gene; cancer; tumour progression; diagnose;
KW hyperproliferative cell growth; prostatic disorder; treatment;
KW metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.
XX
OS Homo sapiens.
XX
PN WO9964594-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-US13181.
XX
PR 11-JUN-1998; 98US-0088877.
PR 09-JUN-1999; 99US-0088877.
XX
PA (CHIR) CHIRON CORP.
XX
PI Astel JH, Carroll E, Endege WO, Ford DM, Monahan JE, Schlegel R;
PI Steinmann KE, Zhang J;
XX
DR WPI; 2000-116541/10.
XX
PT New isolated prostate cancer specific nucleic acids, used to develop
PT products for the diagnosis and treatment of cancer -
XX
PS Claim 2; Page 107; 212pp; English.
XX
CC This sequence represents a prostate cancer specific nucleic acid
CC sequence. The invention relates to a method for diagnosing cancer,
CC tumour progression, hyperproliferative cell growth or accompanying
CC biological and physical manifestations. The method involves contacting
CC the biological sample with a probe that comprises a sequence capable of
CC hybridising to any of the 339 nucleotide sequences given in the
CC specification (see AAZ97140-Z97478) and detecting duplex formation. The
CC products and methods of the invention can be used for the diagnosis,
CC prognosis, and treatment of cancer, tumour progression,
CC hyperproliferative cell growth, and accompanying physical and biological
CC manifestations. They can be used particularly for prostatic disorders
CC such as metastatic prostate cancer, localised prostate cancer, or benign
CC prostate hyperplasia (BPH).
XX
SQ Sequence 1024 BP; 272 A; 243 C; 275 G; 207 T; 27 other;

Query Match 74.2%; Score 399.4; DB 21; Length 1024;
Best Local Similarity 98.5%; Pred. No. 7.3e-106;

Matches 403; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy    130 ACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGG 189
          | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db    100 aggtacagcctcctgaaatgattgggcctatgcggcccgagcagttcagtgatgaagcgg 159

Qy    190 AACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTC 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 aaccagcaacacctgaagaaggggaaccagcaactcaacgtcaggatcctgcagctgctc 219

Qy    250 AGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCC 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 aggagggagaggatgagggagcatctgcaggtcaagggccgaagcctgaagctcatagcc 279

Qy    310 AGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGA 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 aggaacagggtcacccacagactgggtgtgagtgtgaagatggtcctgatgggcaggaga 339

Qy    370 TGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGT 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 tggacccgccaaatccagaggaggtgaaaacgcctgaagaaggtgaaaagcaatcacagt 399

Qy    430 GTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTA 489
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 gttaaaagaaggcacgttgaaatgatgcaggctgctcctatgttggaaatttgttcatta 459

Qy    490 AAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAAA 538
          |||||||||||||||||||||||||||||| ||| ||||||||||||
Db    460 aaattctcccaataaagctttacagccttctgtaaaaaaaaaaaaaaaa 508
```

RESULT 13
AAZ97216/c
ID AAZ97216 standard; cDNA; 1024 BP.
XX
AC AAZ97216;
XX
DT 18-APR-2000 (first entry)
XX
DE Human prostate cancer differentially expressed gene #77.
XX
KW Prostate cancer specific gene; cancer; tumour progression; diagnose;
KW hyperproliferative cell growth; prostatic disorder; treatment;
KW metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.
XX
OS Homo sapiens.
XX
PN WO9964594-A2.
XX
PD 16-DEC-1999.
XX
PF 10-JUN-1999; 99WO-US13181.
XX
PR 11-JUN-1998; 98US-0088877.
PR 09-JUN-1999; 99US-0088877.
XX
PA (CHIR) CHIRON CORP.
XX
PI Astel JH, Carroll E, Endege WO, Ford DM, Monahan JE, Schlegel R;
PI Steinmann KE, Zhang J;
XX
DR WPI; 2000-116541/10.
XX
PT New isolated prostate cancer specific nucleic acids, used to develop
PT products for the diagnosis and treatment of cancer -
XX
PS Claim 2; Page 107; 212pp; English.
XX
CC This sequence represents a prostate cancer specific nucleic acid
CC sequence. The invention relates to a method for diagnosing cancer,
CC tumour progression, hyperproliferative cell growth or accompanying
CC biological and physical manifestations. The method involves contacting
CC the biological sample with a probe that comprises a sequence capable of

```
CC   hybridising to any of the 339 nucleotide sequences given in the
CC   specification (see AAZ97140-Z97478) and detecting duplex formation. The
CC   products and methods of the invention can be used for the diagnosis,
CC   prognosis, and treatment of cancer, tumour progression,
CC   hyperproliferative cell growth, and accompanying physical and biological
CC   manifestations. They can be used particularly for prostatic disorders
CC   such as metastatic prostate cancer, localised prostate cancer, or benign
CC   prostate hyperplasia (BPH).
XX
SQ   Sequence 1024 BP; 187 A; 290 C; 232 G; 270 T; 45 other;


  Query Match             73.9%;  Score 397.4;  DB 21;  Length 1024;
  Best Local Similarity   98.0%;  Pred. No. 2.8e-105;
  Matches  401;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy     130 ACGTAGAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGG 189
           | ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     521 AGGTNCAGCCTCCTGAAATGATTGGGCCTATGCGGCCCGAGCAGTTCAGTGATGAAGCGG 462

Qy     190 AACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTC 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     461 AACCAGCAACACCTGAAGAAGGGGAACCAGCAACTCAACGTCAGGATCCTGCAGCTGCTC 402

Qy     250 AGGAGGGAGAGGATGAGGGAGCATCTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCC 309
           |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     401 AGGAGGGAGAGGATGAGGGAGCATNTGCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCC 342

Qy     310 AGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGA 369
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 AGGAACAGGGTCACCCACAGACTGGGTGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGA 282

Qy     370 TGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGT 429
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 TGGACCCGCCAAATCCAGAGGAGGTGAAAACGCCTGAAGAAGGTGAAAAGCAATCACAGT 222

Qy     430 GTTAAAAGAAGACACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTA 489
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 GTTAAAAGAAGGCACGTTGAAATGATGCAGGCTGCTCCTATGTTGGAAATTTGTTCATTA 162

Qy     490 AAATTCTCCCAATAAAGCTTTACAGCCTTCTGCAAAGAAAAAAAAAAAA 538
           |||||||||||||||||||||||||||||| ||| ||||||||||||
Db     161 AAATTCTCCCAATAAAGCTTTACAGCCTTCTGTAAAAAAAAAAAAAAAA 113


RESULT  14
AAT31861
ID   AAT31861 standard; cDNA; 646 BP.
XX
AC   AAT31861;
XX
DT   07-MAR-1997  (first entry)
XX
DE   GAGE tumour rejection antigen precursor gene.
XX
KW   GAGE gene; tumour rejection antigen; TRA; precursor; TRAP; human;
KW   melanoma cell line; HLA-A29 molecule; binding peptide; diagnosis;
KW   cancer; typing assay; skin graft; organ transplant; MHC;
KW   major histocompatibility complex; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             49..469
FT                   /*tag=  a
FT                   /note= "putative open reading frame"
XX
PN   WO9621673-A1.
XX
PD   18-JUL-1996.
XX
PF   11-JAN-1996;   96WO-US00381.
XX
PR   21-SEP-1995;   95US-0531662.
PR   10-JAN-1995;   95US-0370648.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B,  Van Der Bruggen P;
XX
DR   WPI; 1996-342237/34.
XX
PT   Isolated tumour rejection antigen genes and peptide(s) - used to
PT   develop prods. for diagnosis and treatment of abnormal cellular
PT   conditions, e.g. cancers
XX
PS   Example 6; Page 43; 61pp; English.
XX
CC   The present sequence is GAGE tumour rejection antigen (TRA)
CC   precursor (TRAP) gene, derived from the human melanoma cell line
CC   MZ2-MEL (see PCT/US92/04354). It encodes peptides that are
CC   processed into human MHC HLA-A29 mol. binding peptides, i.e. TRA.
CC   The TRA peptides can be used in the diagnosis of pathological
CC   conditions, e.g. cancer, and for carrying out HLA-typing assays,
CC   e.g. for skin grafts or organ transplants.
XX
SQ   Sequence 646 BP; 188 A; 140 C; 174 G; 144 T; 0 other;


  Query Match             69.6%;  Score 374.4;  DB 17;  Length 646;
  Best Local Similarity   99.7%;  Pred. No. 1.1e-98;
  Matches  375;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      36 CTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGA 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ctgccgtccggactcttttcctctactgagattcatctgtgtgaaatatgagttggcga 60

Qy      96 GGAAGATCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGG 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ggaagatcgacctatcggcctagaccaagacgctacgtagagcctcctgaaatgattggg 120

Qy     156 CCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAA 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cctatgcggcccgagcagttcagtgatgaagtggaaccagcaacacctgaagaaggggaa 180

Qy     216 CCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCT 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ccagcaactcaacgtcaggatcctgcagctgctcaggagggagaggatgagggagcatct 240

Qy     276 GCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGG 335
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     241 gcaggtcaagggccgaagcctgaagctgatagccaggaacagggtcacccacagactggg 300

Qy     336 TGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTG 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgtgagtgtgaagatggtcctgatgggcaggagatggacccgccaaatccagaggaggtg 360

Qy     396 AAAACGCCTGAAGAAG 411
           ||||||||||||||||
Db     361 aaaacgcctgaagaag 376


RESULT  15
AAX90518
ID   AAX90518 standard; cDNA; 646 BP.
XX
AC   AAX90518;
XX
DT   30-SEP-1999  (first entry)
XX
DE   GAGE tumour rejection antigen clone nucleotide sequence.
XX
KW   Human leukocyte antigen; HLA-A29; tumour rejection antigen;
KW   detection; therapy; pathological condition; cancer; CTL;
KW   cytolytic T lymphocyte; GAGE; PCR primer; ss.
XX
```

OS Homo sapiens.
XX
PN WO9937665-A1.
XX
PD 29-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00775.
XX
PR 23-JAN-1998; 98US-0012818.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;
XX
DR WPI; 1999-469111/39.
XX
PT New isolated peptides which bind to HLA-A29 molecules, which are
PT tumour rejection antigens used for detection and therapy of
PT pathological conditions, e.g. cancer
XX
PS Example 9; Page 38; 62pp; English.
XX
CC The present invention describes peptides which bind to human leukocyte
CC antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into
CC tumour rejection antigens. They can be used for detecting cytolytic T
CC lymphocytes (CTLs) in pathological conditions such as cancer and in
CC HLA-typing assays. Complexes of HLA-29 molecules and the peptides can
CC be used for stimulating CTLs in vivo. The present sequence represents
CC a GAGE tumour rejection antigen clone, from an example from the present
CC invention.
XX
SQ Sequence 646 BP; 188 A; 140 C; 174 G; 144 T; 0 other;

Query Match 69.6%; Score 374.4; DB 20; Length 646;
Best Local Similarity 99.7%; Pred. No. 1.1e-98;
Matches 375; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      36 CTGCCGTCCGGACTCTTTTTCCTCTACTGAGATTCATCTGTGTGAAATATGAGTTGGCGA 95
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ctgccgtccggactcttttcctctactgagattcatctgtgtgaaatatgagttggcga 60

Qy      96 GGAAGATCGACCTATCGGCCTAGACCAAGACGCTACGTAGAGCCTCCTGAAATGATTGGG 155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ggaagatcgacctatcggcctagaccaagacgctacgtagagcctcctgaaatgattggg 120

Qy     156 CCTATGCGGCCCGAGCAGTTCAGTGATGAAGTGGAACCAGCAACACCTGAAGAAGGGGAA 215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cctatgcggcccgagcagttcagtgatgaagtggaaccagcaacacctgaagaaggggaa 180

Qy     216 CCAGCAACTCAACGTCAGGATCCTGCAGCTGCTCAGGAGGGAGAGGATGAGGGAGCATCT 275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ccagcaactcaacgtcaggatcctgcagctgctcaggagggagaggatgagggagcatct 240

Qy     276 GCAGGTCAAGGGCCGAAGCCTGAAGCTCATAGCCAGGAACAGGGTCACCCACAGACTGGG 335
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     241 gcaggtcaagggccgaagcctgaagctgatagccaggaacagggtcacccacagactggg 300

Qy     336 TGTGAGTGTGAAGATGGTCCTGATGGGCAGGAGATGGACCCGCCAAATCCAGAGGAGGTG 395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 tgtgagtgtgaagatggtcctgatgggcaggagatggacccgccaaatccagaggaggtg 360

Qy     396 AAAACGCCTGAAGAAG 411
           ||||||||||||||||
Db     361 aaaacgcctgaagaag 376
```

Search completed: July 1, 2002, 06:44:03
Job time: 1938 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:29:18 ; Search time 14.38 Seconds
(without alignments)
775.130 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	119	18.4	1300	2	T03166	probable immediate
2	110	17.0	139	2	S70010	glutamate/proline-
3	110	17.0	221	2	S70009	glutamate/proline-
4	106	16.4	301	2	E29149	proline-rich prote
5	105.5	16.3	260	2	S22373	proline-rich prote
6	104	16.1	396	1	A58938	surface protein rh
7	103.5	16.0	171	2	A27307	proline-rich phosp
8	102.5	15.9	815	2	B30843	glutenin high mole
9	102.5	15.9	815	2	JN0689	glutenin, high-mol
10	102.5	15.9	830	2	S15720	glutenin high mole
11	102.5	15.9	848	2	S02262	glutenin high mole
12	102	15.8	652	2	PC4014	calcium binding 14
13	102	15.8	1560	2	T30282	calcium-binding pr
14	101.5	15.7	300	2	S19560	proline-rich prote
15	101	15.6	754	2	B97586	hypothetical prote
16	101	15.6	754	2	AC2807	OmpA family protei
17	100.5	15.6	605	1	FWSYBA	beta-conglycinin a
18	100.5	15.6	847	2	F96531	hypothetical prote
19	100	15.5	1881	2	H95076	zinc metalloprotei
20	99.5	15.4	309	2	S10889	proline-rich prote
21	99.5	15.4	548	1	QFPGL	neurofilament trip
22	99.5	15.4	605	2	S20007	beta-conglycinin a
23	99.5	15.4	1110	2	I51116	NF-180 - sea lampr
24	98	15.2	212	2	B36298	proline-rich prote
25	98	15.2	838	1	EEWTHW	glutenin, high mol
26	98	15.2	1606	2	T34073	paranemin - chicke
27	98	15.2	2649	2	T51023	hypothetical prote
28	97.5	15.1	166	2	B25372	salivary proline-r
29	97.5	15.1	3436	2	S55659	tegument protein 6
30	97	15.0	204	2	A39066	proline-rich prote
31	96.5	14.9	229	1	W4WL8	E4 protein - human
32	96.5	14.9	617	2	S42719	actin-binding prot
33	96	14.9	543	1	QFMSL	neurofilament trip
34	96	14.9	554	2	JW0094	neurofilament prot
35	96	14.9	1876	2	E97944	zinc metalloprotei
36	95.5	14.8	544	2	S07144	neurofilament trip
37	95.5	14.8	634	1	GERTX1	matrix glycoprotei
38	94.5	14.6	220	2	A36298	proline-rich prote
39	94.5	14.6	530	2	T48627	hypothetical prote
40	94	14.6	183	2	T20980	hypothetical prote
41	94	14.6	1263	2	T13805	spalt-related prot
42	93.5	14.5	245	1	W4WL5	E4 protein - human
43	93.5	14.5	310	1	PIHUSD	salivary proline-r
44	93.5	14.5	325	2	T16324	hypothetical prote
45	93	14.4	206	1	PIRT3	acidic proline-ric

ALIGNMENTS

RESULT 1
T03166
probable immediate early protein - alcelaphine herpesvirus 1
C;Species: alcelaphine herpesvirus 1
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
C;Accession: T03166
R;Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997
A;Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A;Reference number: Z14840; MUID:97404659
A;Accession: T03166
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1300 <ENS>
A;Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAC58118.1; PID:g2338034
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal hom

Query Match 18.4%; Score 119; DB 2; Length 1300;
Best Local Similarity 35.2%; Pred. No. 0.18;
Matches 45; Conservative 7; Mismatches 44; Indels 32; Gaps 7;

```
Qy      17 VEPPEMIGPMRPEQFSDEVEPATPE--EGE-------PATQRQDPAAAQEG--------E 59
           :| ||  ||  ||    |  | ||  |||       |     |   ||||        |
Db     527 LEGPEGEGPEGPEGPEGE-GPEGPEGPEGEGPEGPEGPERDSPDGPGAQEGPEGPEGPEE 585

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDG------PDGQEMDPPN-------PEEVKT 106
           |||    :||: |     :|    | |  :|      |:| | | |:       || |
Db     586 DEGPEGPEGPEGEGPEGPEGEGPEGLEGPEGDEGPEEPEGPEGDSPDGPGAQEVPEGPKG 645

Qy     107 PEEGEKQS 114
           | ||| ||
Db     646 P-EGECQS 652
```

RESULT 2
S70010
glutamate/proline-rich protein (clone BB.1.4.1) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C;Accession: S70010
R;Geertman, R.; McMahon, A.; Sabban, E.L.
Biochim. Biophys. Acta 1306, 147-152, 1996
A;Title: Cloning and characterization of cDNAs for novel proteins with glutamic ac
A;Reference number: S70009; MUID:96221285
A;Accession: S70010
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-139 <GEE>
A;Cross-references: EMBL:U40628; NID:g1184695; PIDN:AAB05668.1; PID:g1184696

Query Match 17.0%; Score 110; DB 2; Length 139;
Best Local Similarity 31.8%; Pred. No. 0.11;
Matches 34; Conservative 15; Mismatches 36; Indels 22; Gaps 5;

```
Qy      15 RYVEPPEMIGPMRPEQFSDEVEPATPEEGE---PATQRQDPAAAQEGEDEGASAGQGPKP 71
           |  |  |: ||    : |: |:| |||| |    ||:  :|   || | |       |:|
Db      47 RASEEEEIGGP----EDSEAVQPRTPEEPELEAEATKGDEPDLEQEMEPE-------PEP 95

Qy      72 EAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTP--EEGEKQSQC 116
           |   : :  |:   | |  |:      | ||  : |  :||::   |
Db      96 EVEPEPEPEPEPEPEPEPEPE------PEPEPEREPDFQEGDESEGC 136
```

RESULT 3
S70009
glutamate/proline-rich protein (clone 53.1.1.1) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C;Accession: S70009
R;Geertman, R.; McMahon, A.; Sabban, E.L.
Biochim. Biophys. Acta 1306, 147-152, 1996
A;Title: Cloning and characterization of cDNAs for novel proteins with glutamic acid-prc
A;Reference number: S70009; MUID:96221285
A;Accession: S70009
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-221 <GEE>
A;Cross-references: EMBL:U40627; NID:g1184693; PIDN:AAB05667.1; PID:g1184694

Query Match 17.0%; Score 110; DB 2; Length 221;
Best Local Similarity 31.8%; Pred. No. 0.16;
Matches 34; Conservative 15; Mismatches 36; Indels 22; Gaps 5;

```
Qy      15 RYVEPPEMIGPMRPEQFSDEVEPATPEEGE---PATQRQDPAAAQEGEDEGASAGQGPKP 71
           |  |  |: ||    : |: |:| |||| |    ||:  :|   || | |       |:|
Db     129 RASEEEEIGGP----EDSEAVQPRTPEEPELEAEATKGDEPDLEQEMEPE-------PEP 177

Qy      72 EAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTP--EEGEKQSQC 116
           |   : :  |:   | |  |:      | ||  : |  :||::   |
Db     178 EVEPEPEPEPEPEPEPEPEPE------PEPEPEREPDFQEGDESEGC 218
```

RESULT 4
E29149
proline-rich protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 20-Aug-1999
C;Accession: E29149
R;Clements, S.; Mehansho, H.; Carlson, D.M.
J. Biol. Chem. 260, 13471-13477, 1985
A;Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence
A;Reference number: A92501; MUID:86033799
A;Contents: Clone pUMP125
A;Accession: E29149
A;Molecule type: mRNA
A;Residues: 1-301 <CLE>
A;Cross-references: GB:M11897; NID:g200540; PIDN:AAA40001.1; PID:g200541
C;Superfamily: proline-rich protein

Query Match 16.4%; Score 106; DB 2; Length 301;
Best Local Similarity 29.3%; Pred. No. 0.43;
Matches 39; Conservative 10; Mismatches 54; Indels 30; Gaps 7;

```
Qy       4 RGRSTYR-PRPRRYVEPPEMIGP-MRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDE 61
           :| |  | |:|     ||   ||  || |  ::  |  |  | |  :   |   |    :
Db     121 QGSSQQRPPQPGNQQGPPPQGGPQQRPPQPGNQQGP--PPPGGPQQRPPQPGGNQGPPPQ 178

Qy      62 GASAG-------QGPKPEAHSQ--------EQGHPQTGCECEDGPDG------QEMDPPN 100
           |   |       ||| |:   |         :|| || |      || |      |:  ||
Db     179 GGPQGPPRPGNQQGPPPQGGPQQRPTQPGNQQGPPQQG-----GPQGPPRPGNQQCPPPQ 233
```

Qy 101 PEEVKTPEEGEKQ 113
| | :|
Db 234 GGPQGPPRPGNQQ 246

RESULT 5
S22373
proline-rich protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C;Accession: S22373
R;Layfield, R.; Bannister, A.J.; Pierce, E.J.; McDonald, C.J.
Eur. J. Biochem. 204, 591-597, 1992
A;Title: cDNA clones for mouse parotid proline-rich proteins. mRNA regulation by
A;Reference number: S22373; MUID:92174915
A;Accession: S22373
A;Molecule type: mRNA
A;Residues: 1-260 <LAY>
A;Cross-references: EMBL:X63004; NID:g53798; PIDN:CAA44733.1; PID:g53799
C;Superfamily: proline-rich protein

Query Match 16.3%; Score 105.5; DB 2; Length 260;
Best Local Similarity 29.5%; Pred. No. 0.4;
Matches 31; Conservative 15; Mismatches 50; Indels 9; Gaps 4;

```
Qy      11 PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPK 70
           |||      ||   ||  | :  ::  |  | :| |  |:: |   |  | |  :|     ||:
Db     144 PRPGNQQGPPPQGGPQGPSRPGNQQGP--PPQGGP--QQRPP---QPGNQQGPPPQGGPQ 196

Qy      71 --PEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
             |:  :|:   || | :    |  |:  ||     : |: |  |
Db     197 GGPQPGNQQGPPPQGGPQGPPRPGNQQGPPPQGGPQRPPQPGNHQ 241
```

RESULT 6
A58938
surface protein rhoptry ROP1 precursor - Toxoplasma gondii
C;Species: Toxoplasma gondii
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A58938; A45644; S37697
R;Boothroyd, J.C.
submitted to GenBank, July 1995
A;Reference number: A58938
A;Accession: A58938
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-396 <BOO>
A;Cross-references: GB:M71274; NID:g897822; PIDN:AAA69859.1; PID:g897823
A;Note: revision to sequence reported in A45644
R;Ossorio, P.N.; Schwartzman, J.D.; Boothroyd, J.C.
Mol. Biochem. Parasitol. 50, 1-15, 1992
A;Title: A Toxoplasma gondii rhoptry protein associated with host cell penetratic
A;Reference number: A45644; MUID:92178277
A;Accession: A45644
A;Molecule type: mRNA
A;Residues: 'MACRQLLCSVQNLLFFFLRDIYCTDFDT',1-352,'FPQR',358-364,'R',366,'I',393,'
A;Cross-references: EMBL:M71274; NID:g897822
A;Note: sequence extracted from NCBI backbone (NCBIN:85178, NCBIP:85179)
C;Superfamily: surface protein rhoptry
C;Keywords: surface antigen
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-396/Product: surface protein rhoptry #status predicted <MAT>

Query Match 16.1%; Score 104; DB 1; Length 396;
Best Local Similarity 34.8%; Pred. No. 0.77;
Matches 40; Conservative 8; Mismatches 41; Indels 26; Gaps 8;

```
Qy       9 YRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPA-----TQRQDPAAA---QEGE- 59
           | | |  :|| | : ||       |:|    |  || |     |  | || |    |||
Db      75 YEPLP-SFVEAP-VRGP-------DQV----PARGEAALVTEETPAQQPAVALGSAEGEG 121
```

```
Qy      60 --DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGE 111
           | ||         |   |  |: | |    |: ||: ||| :|:  | | |
Db     122 TSTTESASENSEDDDTFHDALQELPEDGLEVRP-PNAQELPPPNVQELPPPTEQE 175
```

RESULT 7
A27307
proline-rich phosphoprotein (gene PRH1, Db allele) - human
N;Alternate names: salivary acidic proline-rich protein
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 29-Aug-1997
C;Accession: A27307
R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.
Am. J. Hum. Genet. 41, 1035-1047, 1987
A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pro
A;Reference number: A27307; MUID:88074309
A;Accession: A27307
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-171 <AZE>
A;Cross-references: EMBL:K03203
C;Genetics:
A;Gene: GDB:PRH1
A;Cross-references: GDB:119515; OMIM:168730
A;Map position: 12p13.2-12p13.2
C;Superfamily: proline-rich protein
C;Keywords: phosphoprotein

Query Match 16.0%; Score 103.5; DB 2; Length 171;
Best Local Similarity 30.9%; Pred. No. 0.39;
Matches 30; Conservative 9; Mismatches 41; Indels 17; Gaps 4;

```
Qy      29 EQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAG---------QGPKPEAHSQEQG 79
           ||| ||     |  |    |:  |:|    :|:|   |          ||| |    : ||
Db      23 EQFLDEERQGPPLGG----QQSQPSAGDGNQDDGPQQGPPQQGGQQQQGP-PPPQGKPQG 77

Qy      80 HPQTGCECEDG---PDGQEMDPPNPEEVKTPEEGEKQ 113
            || | : : |   | |:   ||        | :|  |
Db      78 PPQQGGQQQQGPPPPQGKPQGPPQQGGHPPPPQGRPQ 114
```

RESULT 8
B30843
glutenin high molecular weight chain Ax2 precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
C;Accession: B30843
R;Anderson, O.D.; Greene, F.C.
submitted to GenBank, January 1989
A;Reference number: A94515
A;Accession: B30843
A;Molecule type: DNA
A;Residues: 1-815 <AND>
C;Superfamily: glutenin
C;Keywords: seed; storage protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-815/Product: glutenin Ax2 chain #status predicted <MAT>
F;108-773/Region: glutamine/glycine/proline-rich

Query Match 15.9%; Score 102.5; DB 2; Length 815;
Best Local Similarity 23.7%; Pred. No. 1.9;
Matches 31; Conservative 27; Mismatches 54; Indels 19; Gaps 5;

```
Qy       2 SWRGRSTYRPRPRRYVEPPEMIG-------PMRPEQFSDEVEPATPEEGEPATQRQDPAA 54
           | :|:   : :|  |   |: :|       |  |:|   | :|   :: |   | | |
Db     362 SGQGQQPGQRQPGYYSTSPQQLGQGQPRYYPTSPQQPGQEQQPRQLQQPEQGQQGQQPEQ 421

Qy      55 AQEGE-----DEGASAGQGPKPEAHSQEQG--HPQTGCECEDGPDGQEMDPPNPE---EV 104
             |:|:     ::|   ||| : :   | |   :| :  :   |  |    | :|:   ::
Db     422 GQQGQQQRQGEQGQQPGQGQQGQQPGQGQPGYYPTSPQQSGQGQPGYY--PTSPQQSGQL 479

Qy     105 KTPEEGEKQSQ 115
           : | :|::  |
Db     480 QQPAQGQQPGQ 490
```

RESULT 9
JN0689
glutenin, high-molecular-weight Ax2* chain precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 15-Nov-1996
C;Accession: JN0689
R;Anderson, O.D.; Greene, F.C.
Theor. Appl. Genet. 77, 689-700, 1989
A;Title: The characterization and comparative analysis of high-molecular-weight g
A;Reference number: JN0689
A;Accession: JN0689
A;Molecule type: DNA
A;Residues: 1-815 <AND>
A;Note: the authors translated the codon CTA for residue 11 as Val, CAT for resid
C;Comment: The main wheat storage proteins are divided into two groups. The glute
a families.
C;Genetics:
A;Gene: Ax2*
C;Superfamily: glutenin
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-815/Product: glutenin, high-molecular-weight Ax2* chain #status predicted <M

Query Match 15.9%; Score 102.5; DB 2; Length 815;
Best Local Similarity 23.7%; Pred. No. 1.9;
Matches 31; Conservative 27; Mismatches 54; Indels 19; Gaps 5;

```
Qy       2 SWRGRSTYRPRPRRYVEPPEMIG-------PMRPEQFSDEVEPATPEEGEPATQRQDPAA 54
           | :|:   : :|  |   |: :|       |  |:|   | :|   :: |   | | |
Db     362 SGQGQQPGQRQPGYYSTSPQQLGQGQPRYYPTSPQQPGQEQQPRQLQQPEQGQQGQQPEQ 421

Qy      55 AQEGE-----DEGASAGQGPKPEAHSQEQG--HPQTGCECEDGPDGQEMDPPNPE---EV 104
             |:|:     ::|   ||| : :   | |   :| :  :   |  |    | :|:   ::
Db     422 GQQGQQQRQGEQGQQPGQGQQGQQPGQGQPGYYPTSPQQSGQGQPGYY--PTSPQQSGQL 479

Qy     105 KTPEEGEKQSQ 115
           : | :|::  |
Db     480 QQPAQGQQPGQ 490
```

RESULT 10
S15720
glutenin high molecular weight chain 1Ax1 precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 08-Jun-1994 #sequence_revision 12-May-1995 #text_change 20-Aug-1999
C;Accession: S15720
R;Halford, N.G.; Field, J.M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thomps
submitted to the EMBL Data Library, July 1991
A;Description: Analysis of HMW glutenin subunit encoded by chromosome 1A of bread
A;Reference number: S15720
A;Accession: S15720
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-830 <HAL>
A;Cross-references: EMBL:X61009; NID:g21742; PIDN:CAA43331.1; PID:g21743
C;Superfamily: glutenin

Query Match 15.9%; Score 102.5; DB 2; Length 830;
Best Local Similarity 24.0%; Pred. No. 2;
Matches 31; Conservative 20; Mismatches 63; Indels 15; Gaps 4;

```
Qy       2 SWRGRSTYRPRPRRYVEPPEMIG-------PMRPEQFSDEVEPATPEEGEPATQRQDPAA 54
           | :|:   : :|  |   |: :|       |  |:|   | :|   :: |   | | |
Db     368 SGQGQQPGQRQPGYYSTSPQQLGQGQPRYYPTSPQQPGQEQQPRQLQQPEQGQQGQQPEQ 427

Qy      55 AQEGEDEG-ASAGQGPKPEAHSQEQGHPQTG---CECEDGPDGQ----EMDPPNPEEVKT 106
             |:|:  |    || |      |: |  | |      :    ||       |    :::
```

```
Db    428 GQQGQQPGQGEQGQQPGQGQQGQQPGQGQPGYYPTSPQQSGQGQPGYYPTSPQQSGQLQQ 487

Qy    107 PEEGEKQSQ 115
          | :|::  |
Db    488 PAQGQQPGQ 496
```

RESULT 11
S02262
glutenin high molecular weight chain Dx5 - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Feb-1994
C;Accession: S02262
R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,
Nucleic Acids Res. 17, 461-462, 1989
A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D
A;Reference number: S02262; MUID:89098419
A;Accession: S02262
A;Molecule type: DNA
A;Residues: 1-848 <AND>
A;Cross-references: EMBL:X12928
C;Genetics:
A;Gene: Glu-D1-1b
C;Superfamily: glutenin

```
  Query Match             15.9%;  Score 102.5;  DB 2;  Length 848;
  Best Local Similarity   23.4%;  Pred. No. 2;
  Matches   33;  Conservative   24;  Mismatches   53;  Indels   31;  Gaps    6;

Qy      4 RGRSTYRP----------RPRRYVEPPE------MIGPMRPEQFS---DEVEPATPEEGE 44
          :|:  | |          :| :: :| :         |::| |      |   |  | :|:
Db    491 QGQPGYYPTSPQQSGQGQQPGQWQQPGQGQPGYYPTSPLQPGQGQPGYDPTSPQQPGQGQ 550

Qy     45 PATQRQDPAAAQEGE--------DEGASAGQGPKPE--AHSQEQGHPQTGCECEDGPDGQ 94
            | | ||  |:|:          : |   || :|      |: |  | | :   |  ||
Db    551 QPGQLQQPAQGQQGQQLAQGQQGQQPAQVQQGQRPAQGQQGQQPGQGQQGQQLGQGQQGQ 610

Qy     95 EMDPPNPEEVKTPEEGEKQSQ 115
          :  |   :: : | :|::  |
Db    611 Q--PGQGQQGQQPAQGQQGQQ 629
```

RESULT 12
PC4014
calcium binding 140k protein - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 21-Jul-1995 #sequence_revision 31-Mar-1997 #text_change 05-Nov-1999
C;Accession: PC4014
R;Naved, A.F.; Ozawa, M.; Yu, S.; Miyauchi, T.; Muramatsu, H.; Muramatsu, T.
Cell Struct. Funct. 20, 133-141, 1995
A;Title: CBP-140, a novel endoplasmic reticulum resident Ca2+-binding protein with a car
A;Reference number: PC4014; MUID:95368740
A;Accession: PC4014
A;Molecule type: mRNA
A;Residues: 1-652 <NAV>
A;Cross-references: GB:S78797; NID:g1041986; PIDN:AAB35051.1; PID:g1041987
C;Comment: The sequence 1-80 has high homology to that of 70k heat shock protein.
C;Keywords: calcium binding; endoplasmic reticulum; heat shock; stress-induced protein
F;649-652/Region: endoplasmic reticulum retention signal #status atypical

```
  Query Match             15.8%;  Score 102;  DB 2;  Length 652;
  Best Local Similarity   34.1%;  Pred. No. 1.7;
  Matches   31;  Conservative    8;  Mismatches   36;  Indels   16;  Gaps    4;

Qy     25 PMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE--DEGASAGQGPKPEAHSQEQGHPQ 82
          |    :  :| ||   |  :|        ||:||| ||  |   | ||||   |:|
Db    270 PAEQGELKEEAEPPAEETSQPPPSEPKGDAAREGEKPDEKES---GDKPEAQPNEKG--- 323

Qy     83 TGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
             : ||:|     | ||| |  :   ||
Db    324 ----QAGPEGA---APAPEEDKKLKPARKQ 346
```

RESULT 13
T30282
calcium-binding protein - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C;Accession: T30282
R;Soltysik-Espanola, M.; Klinzing, D.C.; Pfarr, K.; Burke, R.D.; Ernst, S.G.
Dev. Biol. 165, 73-85, 1994
A;Title: Endo16, a large multidomain protein found on the surface and ECM of enc
A;Reference number: Z20805; MUID:94374583
A;Accession: T30282
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1560 <SOL>
A;Cross-references: EMBL:L34680; NID:g511893; PID:g511894; PIDN:AAA30047.1
C;Genetics:
A;Note: endo16

```
  Query Match             15.8%;  Score 102;  DB 2;  Length 1560;
  Best Local Similarity   32.0%;  Pred. No. 3.8;
  Matches   32;  Conservative   13;  Mismatches   35;  Indels   20;  Gaps    5;

Qy     34 EVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECE----- 88
          ||: :  |  |   : |||    ||| :| |  :| :||    ||| ||   | |
Db   1362 EVKTSNQEGSEVQEESQDPQEQMEGEGQG-SESEGEEPEEPEGEQG-PQEQVESEGEENE 1419

Qy     89 ------DGP------DGQEMDPPNPE-EVKTPEEGEKQSQ 115
                : |      :||: :   || | : ||| | : :
Db   1420 SEEEEVEEPQGVMESEGQQSESEEPEFEGEEPEESEGEQE 1459
```

RESULT 14
S19560
proline-rich protein MP4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
C;Accession: S19560; S22570
R;Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.
Eur. J. Biochem. 202, 969-974, 1991
A;Title: Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional
A;Reference number: S19560; MUID:92111548
A;Accession: S19560
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-300 <ROB>
A;Cross-references: GB:X58438; NID:g53181; PIDN:CAA41344.1; PID:g53182
R;Roberts, S.G.E.; Layfield, R.; McDonald, C.J.
Nucleic Acids Res. 19, 5205-5211, 1991
A;Title: The mouse proline-rich protein MP6 promoter binds isoprenaline-inducibl
A;Reference number: S22570; MUID:92020206
A;Accession: S22570
A;Molecule type: DNA
A;Residues: 1-14 <RO2>
A;Cross-references: EMBL:X61126
C;Superfamily: proline-rich protein

```
  Query Match             15.7%;  Score 101.5;  DB 2;  Length 300;
  Best Local Similarity   28.8%;  Pred. No. 0.91;
  Matches   38;  Conservative   10;  Mismatches   55;  Indels   29;  Gaps    7;

Qy      4 RGRSTYR-PRPRRYVEPPEMIGP-MRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDE 61
          :| |  | |:|     ||   ||  || |  ::  |  |  | |  :   |  | | :
Db    121 QGSSQQRPPQPGNQQGPPPQGGPQQRPPQPGNQQGP--PPPGGPQQRPPQPGGNQGGPPQ 178

Qy     62 GA-------SAGQGPKPEAHSQ--------EQGHPQTGCECEDGPD-----GQEMDPPNP 101
          |           ||| |:   |         :|| || |     ||      | :  ||
Db    179 GGPHPPPRPGNQQGPPPQGGPQQRPTQPGNQQGPPQQG-----GPQAPPRPGNQQGPPPQ 233

Qy    102 EEVKTPEEGEKQ 113
```

```
                |  | :|
Db    234 GPQGPPRTGNQQ 245


RESULT  15
B97586
hypothetical protein AGR_C_3445 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: B97586
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: B97586
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-754 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87643.1; PID:g15156995; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3445
A;Map position: circular chromosome


  Query Match             15.6%;  Score 101;  DB 2;  Length 754;
  Best Local Similarity   27.7%;  Pred. No. 2.3;
  Matches   36;  Conservative   15;  Mismatches   47;  Indels   32;  Gaps    5;

Qy      10 RPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQ-- 67
           |||  :  | |   |  |||:   | |||     ||| ::| |||  |  :: |   :
Db     182 RPRKPQKAEQPAGEGEQRPERPRKEKEPAK----EPAAEQQ-PAARPENAEQPAKPREPA 236

Qy      68 -GPKPEAHSQEQGHPQTGCECEDGPDGQEMD--------------------PPNPEEVK 105
            | ||:    |:  |:   |  : |  ::                     || ||
Db     237 PGKKPQV---EEKAPEQKAEPAEKPVPEKKPAAPEPAAKEAPVPTEAPTPARPPAPEAQP 293

Qy     106 TPEEGEKQSQ 115
           |  | : |:
Db     294 NPAPGRQPSE 303


Search completed: July  1, 2002, 06:33:05
Job time: 227 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:32:39 ; Search time 12.4 Seconds
(without alignments)
362.215 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	646	100.0	116	1	GGE2_HUMAN	Q13066 homo sapien
2	625.5	96.8	117	1	GGE7_HUMAN	O76087 homo sapien
3	616.5	95.4	117	1	GGE4_HUMAN	Q13068 homo sapien
4	612.5	94.8	117	1	GGE5_HUMAN	Q13069 homo sapien
5	608.5	94.2	117	1	GGE6_HUMAN	Q13070 homo sapien
6	598	92.6	138	1	GGE1_HUMAN	Q13065 homo sapien
7	589.5	91.3	118	1	GGE3_HUMAN	Q13067 homo sapien
8	266	41.2	146	1	GGB1_HUMAN	O75459 homo sapien
9	120	18.6	102	1	GGC1_HUMAN	O60829 homo sapien
10	100.5	15.6	504	1	TPM5_DROME	P49456 drosophila
11	100.5	15.6	605	1	GLCA_SOYBN	P13916 glycine max
12	99.5	15.4	548	1	NFL_PIG	P02547 sus scrofa
13	99	15.3	839	1	GLT5_WHEAT	P10388 triticum ae
14	98	15.2	838	1	GLT4_WHEAT	P08489 triticum ae
15	97.5	15.1	166	1	PRPC_HUMAN	P02810 homo sapien
16	97.5	15.1	235	1	BAR6_CHITE	P08726 chironomus
17	97.5	15.1	541	1	NFL_RAT	P19527 rattus norv
18	97	15.0	261	1	PRP2_MOUSE	P05142 mus musculu
19	96.5	14.9	229	1	VE4_HPV08	P06425 human papil
20	96.5	14.9	617	1	ABP1_SACEX	P38479 saccharomyc
21	96	14.9	542	1	NFL_MOUSE	P08551 mus musculu
22	96	14.9	554	1	NFL_BOVIN	P02548 bos taurus
23	95.5	14.8	543	1	NFL_HUMAN	P07196 homo sapien
24	95.5	14.8	634	1	SC1_RAT	P24054 rattus norv
25	94.5	14.6	245	1	VE4_HPV05	P06924 human papil
26	94.5	14.6	247	1	PRP4_HUMAN	P10163 homo sapien
27	94	14.6	234	1	PRPM_HUMAN	P10161 homo sapien
28	93.5	14.5	276	1	PRPL_HUMAN	P10162 homo sapien
29	93.5	14.5	580	1	SYN3_HUMAN	O14994 homo sapien
30	93	14.4	174	1	PRPP_HUMAN	P81489 homo sapien
31	93	14.4	206	1	PRP3_RAT	P04474 rattus norv
32	92.5	14.3	183	1	PRPH_MESAU	P06680 mesocricetu
33	92	14.2	466	1	CMGA_RAT	P10354 rattus norv
34	92	14.2	466	1	SX10_MOUSE	Q04888 mus musculu
35	92	14.2	533	1	SX21_MOUSE	O09141 mus musculu
36	92	14.2	595	1	VIE1_MCMVS	P11210 murine cyto
37	91.5	14.2	245	1	VE4_HPV5B	P26550 human papil
38	91.5	14.2	579	1	SYN3_RAT	O70441 rattus norv
39	91.5	14.2	601	1	CORO_SCHPO	O13923 schizosacch
40	91.5	14.2	1099	1	NKX1_HUMAN	O60721 homo sapien
41	91.5	14.2	1298	1	ICP4_HSV11	P08392 herpes simp
42	91	14.1	417	1	VGLD_HSVBS	Q08100 bovine herp
43	91	14.1	424	1	MR15_DROME	Q9y0i1 drosophila
44	90.5	14.0	446	1	CMGA_PIG	P04404 sus scrofa
45	90.5	14.0	884	1	ANDR_EULFC	O97776 eulemur ful

ALIGNMENTS

RESULT 1
GGE2_HUMAN
ID GGE2_HUMAN STANDARD; PRT; 116 AA.
AC Q13066;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GAGE-2 protein (G antigen 2).
GN GAGE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Melanoma;
RX MEDLINE=95378788; PubMed=7544395;
RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA Boon T.;
RT "A new family of genes coding for an antigen recognized by autologous
RT cytolytic T lymphocytes on a human melanoma.";
RL J. Exp. Med. 182:689-698(1995).
CC -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC T LYMPHOCYTES.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U19143; AAA82745.1; -.
DR MIM; 604244; -.
KW Multigene family; Antigen.
SQ SEQUENCE 116 AA; 12786 MW; DD305D5CA29AF19A CRC64;

Query Match 100.0%; Score 646; DB 1; Length 116;
Best Local Similarity 100.0%; Pred. No. 3.8e-38;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
```

RESULT 2

```
GGE7_HUMAN
ID   GGE7_HUMAN     STANDARD;      PRT;   117 AA.
AC   O76087;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-7 protein (G antigen 7) (GAGE-8) (AL4).
GN   GAGE7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316329; PubMed=9651357;
RA   Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
RT   "Isolation and characterization of PAGE-1 and GAGE-7: new genes
RT   expressed in the LNCaP prostate cancer progression model that share
RT   homology with melanoma-associated antigens.";
RL   J. Biol. Chem. 273:17618-17625(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   de Backer O.R.Y.;
RT   "Structure, chromosomal location and control of expression of the
RT   GAGE genes.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SOME PROSTATE CANCER TISSUES BUT
CC       NOT IN NORMAL PROSTATE TISSUE.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF058988; AAC25989.1; -.
DR   EMBL; AF055474; AAC33677.1; -.
DR   EMBL; AF055475; AAC33678.1; -.
DR   MIM; 604132; -.
KW   Multigene family.
SQ   SEQUENCE   117 AA;  12978 MW;  365ED8D41361C7AD CRC64;


  Query Match             96.8%;  Score 625.5;  DB 1;  Length 117;
  Best Local Similarity   97.4%;  Pred. No. 9.6e-37;
  Matches  114;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  ||||||:|||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 117


RESULT   3
GGE4_HUMAN
ID   GGE4_HUMAN     STANDARD;      PRT;   117 AA.
AC   Q13068;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-4 protein (G antigen 4).
GN   GAGE4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19145; AAA82747.1; -.
DR   MIM; 604246; -.
KW   Multigene family.
SQ   SEQUENCE   117 AA;  12956 MW;  365ED71B2F9DC7AD CRC64;


  Query Match             95.4%;  Score 616.5;  DB 1;  Length 117;
  Best Local Similarity   96.6%;  Pred. No. 3.9e-36;
  Matches  113;  Conservative    1;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  ||||||:|||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYYWPRPRRYVQPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db      61 DEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 117


RESULT   4
GGE5_HUMAN
ID   GGE5_HUMAN     STANDARD;      PRT;   117 AA.
AC   Q13069;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-5 protein (G antigen 5).
GN   GAGE5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U19146; AAA82748.1; -.
DR   MIM; 604247; -.
KW   Multigene family.
SQ   SEQUENCE   117 AA;  12924 MW;  234A865E2ECDDD06 CRC64;


  Query Match             94.8%;  Score 612.5;  DB 1;  Length 117;
  Best Local Similarity   95.7%;  Pred. No. 7.3e-36;
  Matches  112;  Conservative    2;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  ||||||:|||:||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      61 DEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 117


RESULT   5
GGE6_HUMAN
ID   GGE6_HUMAN     STANDARD;      PRT;   117 AA.
AC   Q13070;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-6 protein (G antigen 6).
GN   GAGE6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19147; AAA82749.1; -.
DR   MIM; 604248; -.
KW   Multigene family.
SQ   SEQUENCE   117 AA;  12892 MW;  234A865E3FCCCD06 CRC64;


  Query Match             94.2%;  Score 608.5;  DB 1;  Length 117;
  Best Local Similarity   94.9%;  Pred. No. 1.4e-35;
  Matches  111;  Conservative    3;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  |||||||:|||:||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||||:||||||||||||||||||||||
Db      61 DEGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEVDPPNPEEVKTPEEGEKQSQC 117


RESULT   6
GGE1_HUMAN
ID   GGE1_HUMAN     STANDARD;      PRT;   138 AA.
AC   Q13065;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-1 protein (G antigen 1) (MZ2-F antigen).
GN   GAGE1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC
CC       T LYMPHOCYTES. COMPLETELY SILENT IN NORMAL ADULT TISSUES, EXCEPT
CC       TESTIS.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19142; AAA82744.1; -.
DR   MIM; 604243; -.
KW   Multigene family; Antigen.
SQ   SEQUENCE   138 AA;  15418 MW;  37B8F3909EC4B3B2 CRC64;


  Query Match             92.6%;  Score 598;  DB 1;  Length 138;
  Best Local Similarity   99.1%;  Pred. No. 8.2e-35;
  Matches  108;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEE 109
           ||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      61 EGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEE 109


RESULT   7
GGE3_HUMAN
ID   GGE3_HUMAN     STANDARD;      PRT;   118 AA.
AC   Q13067;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GAGE-3 protein (G antigen 3).
GN   GAGE3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=95378788; PubMed=7544395;
RA   van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S.,
```

```
RA   Boon T.;
RT   "A new family of genes coding for an antigen recognized by autologous
RT   cytolytic T lymphocytes on a human melanoma.";
RL   J. Exp. Med. 182:689-698(1995).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT
CC       NOT IN NORMAL TISSUES, EXCEPT TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19144; AAA82746.1; -.
DR   MIM; 604245; -.
KW   Multigene family.
SQ   SEQUENCE   118 AA;  12937 MW;  D97EEBB19E735103 CRC64;
```

```
  Query Match             91.3%;  Score 589.5;  DB 1;  Length 118;
  Best Local Similarity   94.7%;  Pred. No. 2.7e-34;
  Matches  108;  Conservative    3;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       4 RGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEG 62
           ||:|||  |||||||:|||:||||||||||||||||||||||||||||||||||||||||
Db       5 RGKSTYYWPRPRRYVQPPEVIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEG 64

Qy      63 ASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      65 ASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 118
```

```
RESULT   8
GGB1_HUMAN
ID   GGB1_HUMAN     STANDARD;      PRT;   146 AA.
AC   O75459;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G antigen family B 1 protein (Prostate-associated gene protein 1)
DE   (PAGE-1) (GAGE-9) (AL5).
GN   GAGEB1 OR PAGE1 OR GAGE9.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98316329; PubMed=9651357;
RA   Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
RT   "Isolation and characterization of PAGE-1 and GAGE-7: new genes
RT   expressed in the LNCaP prostate cancer progression model that share
RT   homology with melanoma-associated antigens.";
RL   J. Biol. Chem. 273:17618-17625(1998).
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;
CC       EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE
CC       PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL
CC       PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF058989; AAC25990.1; -.
DR   MIM; 300288; -.
KW   Multigene family.
SQ   SEQUENCE   146 AA;  16134 MW;  91ABA1E3B498DCA2 CRC64;
```

```
  Query Match             41.2%;  Score 266;  DB 1;  Length 146;
  Best Local Similarity   42.0%;  Pred. No. 3.2e-12;
  Matches   66;  Conservative    7;  Mismatches   30;  Indels   54;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           | :  |  || ||  |||          |: |||     |:| |  || ||   |:| ||
Db       1 MGFLRRLIYRRRPMIYVE--------SSEESSDE----QPDEVESPTQSQDSTPAEERED 48

Qy      61 EGASAGQ----------------------------------------------GPKPEAHSQEQ 78
           |||||  |                                              ||:||| ||||
Db      49 EGASAAQGQEPEADSQELVQPKTGCEPGDGPDTKRVCLRNEEQMKLPAEGPEPEADSQEQ 108

Qy      79 GHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
           ||:||||  |||| ||:  ||||||||||| | |||
Db     109 VHPKTGCERGDGPDVQELGLPNPEEVKTPEEDEGQSQ 145
```

```
RESULT   9
GGC1_HUMAN
ID   GGC1_HUMAN     STANDARD;      PRT;   102 AA.
AC   O60829;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   G antigen family C 1 protein (Prostate-associated gene protein 4)
DE   (PAGE-4) (PAGE-1) (JM27) (GAGE-9).
GN   GAGEC1 OR PAGE4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98393718; PubMed=9724777;
RA   Brinkmann U., Vasmatzis G., Lee B., Yerushalmi N., Essand M.,
RA   Pastan I.;
RT   "PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in
RT   normal and neoplastic prostate, testis, and uterus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strom T.M., Nyakatura G., Hellebrand H., Drescher B., Rosenthal A.,
RA   Meindl A.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: UNKNOWN.
CC   -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND
CC       FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,
CC       UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE CANCER, TESTICULAR
CC       CANCER, AND UTERINE CANCER.
CC   -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF275258; AAF88037.1; -.
DR   EMBL; AJ005894; CAA06751.1; -.
DR   EMBL; AF238380; AAF62541.1; -.
DR   MIM; 300287; -.
SQ   SEQUENCE   102 AA;  11153 MW;  CE5D07AFBF73301B CRC64;
```

```
  Query Match             18.6%;  Score 120;  DB 1;  Length 102;
  Best Local Similarity   34.5%;  Pred. No. 0.02;
```

```
  Matches   39;  Conservative   10;  Mismatches   52;  Indels   12;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           || | ||  | |     | |:::           | |   :: || |  ||     |  |
Db       1 MSARVRSRSRGRGDGQ-EAPDVVA---------FVAPGESQQEEPPTDNQDIEPGQ--ER 48

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           ||    :  | |   ||    :|  |  || | :|  ||||:  || | |: |
Db      49 EGTPPIEERKVEGDCQEMDLEKTRSERGDGSDVKEKTPPNPKHAKTKEAGDGQ 101


RESULT  10
TPM5_DROME
ID   TPM5_DROME     STANDARD;      PRT;   504 AA.
AC   P49456;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Tropomyosin 1, fusion protein 34.
GN   TM1 OR TMII.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89127197; PubMed=2851721;
RA   Hanke P.D., Storti R.V.;
RT   "The Drosophila melanogaster tropomyosin II gene produces multiple
RT   proteins by use of alternative tissue-specific promoters and
RT   alternative splicing.";
RL   Mol. Cell. Biol. 8:3591-3602(1988).
RN   [2]
RP   SEQUENCE OF 1-286 FROM N.A.
RX   MEDLINE=87064486; PubMed=3097506;
RA   Karlik C.C., Fyrberg E.A.;
RT   "Two Drosophila melanogaster tropomyosin genes: structural and
RT   functional aspects.";
RL   Mol. Cell. Biol. 6:1965-1973(1986).
CC   -!- ALTERNATIVE PRODUCTS: DROSOPHILA TROPOMYOSIN 1 GENE CAN PRODUCE
CC       FOUR DIFFERENT ISOFORMS BY ALTERNATIVE PRODUCTS: A MUSCLE FORM,
CC       A NON-MUSCLE FORM, AND TWO FUSION PROTEINS (33 AND 34).
CC   -!- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC       EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC   -!- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN MAINLY BECAUSE
CC       OF ALTERNATIVE EXON USAGE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76208; CAA53801.1; -.
DR   EMBL; L00362; AAA28965.1; -.
DR   EMBL; M12840; AAA28965.1; JOINED.
DR   EMBL; L00355; AAA28965.1; JOINED.
DR   EMBL; L00356; AAA28965.1; JOINED.
DR   EMBL; L00357; AAA28965.1; JOINED.
DR   EMBL; L00358; AAA28965.1; JOINED.
DR   EMBL; L00359; AAA28965.1; JOINED.
DR   FlyBase; FBgn0003721; Tm1.
DR   InterPro; IPR000533; Tropomyosin.
DR   Pfam; PF00261; Tropomyosin; 1.
DR   PRINTS; PR00194; TROPOMYOSIN.
DR   PROSITE; PS00326; TROPOMYOSIN; 1.
KW   Coiled coil; Repeat; Alternative splicing; Multigene family.
FT   DOMAIN       14    275       COILED COIL (POTENTIAL).
FT   DOMAIN      298    504       ALA/PRO-RICH.
FT   CONFLICT    106    114       LGSATAKLS -> SASAIQLAA (IN REF. 2).
FT   CONFLICT    119    119       A -> S (IN REF. 2).
FT   CONFLICT    183    183       A -> AMVEADLERAEERA (IN REF. 2).
FT   CONFLICT    199    199       V -> L (IN REF. 2).
FT   CONFLICT    215    231       NQREEEYKNQIKTLNTR -> TQKEETFETQIKVLDHS
FT                                (IN REF. 2).
SQ   SEQUENCE   504 AA;  53634 MW;  642C5D73F03AAC43 CRC64;


  Query Match             15.6%;  Score 100.5;  DB 1;  Length 504;
  Best Local Similarity   33.3%;  Pred. No. 1.7;
  Matches   31;  Conservative    9;  Mismatches   42;  Indels   11;  Gaps    4;

Qy      15 RYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAH 74
           || :||    |   |:    :| |   | |   : :: ||   ||: || ||:|  | |
Db     291 RYPKPPT---PELTEEEKAALEAAAIAEAEAKARAEELAAL--GEEAGAEAGEGGAP-AE 344

Qy      75 SQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTP 107
             | |    |      | |  |  || : ||
Db     345 GAAPGEPGAATE-----PGVEAPPAEPERIPTP 372


RESULT  11
GLCA_SOYBN
ID   GLCA_SOYBN     STANDARD;      PRT;   605 AA.
AC   P13916;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Beta-conglycinin, alpha chain precursor.
OS   Glycine max (Soybean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX   NCBI_TaxID=3847;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Cotyledon;
RX   MEDLINE=91355860; PubMed=2103438;
RA   Sebastiani F.L., Farrel L.B., Schuler M.A., Beachy R.N.;
RT   "Complete sequence of a cDNA of alpha subunit of soybean beta-
RT   conglycinin.";
RL   Plant Mol. Biol. 15:197-201(1990).
CC   -!- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED
CC       DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A
CC       CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC   -!- SUBUNIT: THE ALPHA'-, ALPHA-, AND BETA-SUBUNITS ASSOCIATE IN
CC       VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC   -!- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONARY MEMBRANE-BOUND
CC       VACUOLAR PROTEIN BODIES.
CC   -!- SIMILARITY: BELONGS TO THE 7S SEED STORAGE PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17698; CAA35691.1; -.
DR   PIR; S14681; FWSYBA.
DR   HSSP; P50477; 1CAU.
DR   InterPro; IPR001113; Seedstore_7s.
DR   Pfam; PF00546; Seedstore_7s; 1.
DR   Pfam; PF02808; Seedstore_7s_C; 1.
KW   Seed storage protein; Signal; Glycoprotein; Multigene family.
FT   SIGNAL       1     22
FT   PROPEP      23     62
FT   CHAIN       63    605       BETA-CONGLYCININ, ALPHA CHAIN.
FT   CARBOHYD    261    261       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   605 AA;  70293 MW;  CBEBA30506BBBC57 CRC64;
```

```
  Query Match             15.6%;  Score 100.5;  DB 1;  Length 605;
  Best Local Similarity   28.1%;  Pred. No. 2;
  Matches   34;  Conservative   12;  Mismatches   40;  Indels   35;  Gaps    5;

Qy      10 RPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQD-------PAAAQEGEDEG 62
           |||||          |  ||:     ||  | | |     |         |   || | |
Db      75 RPRPR----------PQHPER-----EPQQPGEKEEDEDEQPRPIPFPRPQPRQEEEHEQ 119

Qy      63 ASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMD--------PPNPEEVKTPEEGEKQS 114
             : |:     :|:   : | | ||  : :| |        ||: :| :  || | :
Db     120 REEQEWPR-----KEEKRGEKGSEEEDEDEDEEQDERQFPFPRPPHQKEERNEEEDEDEE 174

Qy     115 Q 115
           |
Db     175 Q 175
```

```
RESULT  12
NFL_PIG
ID   NFL_PIG        STANDARD;      PRT;   548 AA.
AC   P02547;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet L protein (68 kDa neurofilament protein)
DE   (Neurofilament light polypeptide) (NF-L).
GN   NEFL.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85154583; PubMed=3920075;
RA   Geisler N., Plessmann U., Weber K.;
RT   "The complete amino acid sequence of the major mammalian
RT   neurofilament protein (NF-L).";
RL   FEBS Lett. 182:475-478(1985).
RN   [2]
RP   SEQUENCE OF 1-82 AND 278-548.
RA   Geisler N., Kaufmann E., Fischer S., Plessmann U., Weber K.;
RT   "Neurofilament architecture combines structural principles of
RT   intermediate filaments with carboxy-terminal extensions increasing
RT   in size between triplet proteins.";
RL   EMBO J. 2:1295-1302(1983).
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC   -!- DOMAIN: THE EXTRA MASS AND HIGH CHARGE DENSITY THAT DISTINGUISH
CC       THE NEUROFILAMENT PROTEINS FROM ALL OTHER INTERMEDIATE FILAMENT
CC       PROTEINS ARE DUE TO THE TAILPIECE EXTENSIONS. THIS REGION MAY FORM
CC       A CHARGED SCAFFOLDING STRUCTURE SUITABLE FOR INTERACTION WITH
CC       OTHER NEURONAL COMPONENTS OR IONS.
CC   -!- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE
CC       NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL
CC       INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 10-NM
CC       FILAMENTS.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR   PIR; A02963; QFPGL.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Glycoprotein.
FT   DOMAIN        1     91       HEAD.
FT   DOMAIN       92    395       ROD.
FT   DOMAIN      396    548       TAIL.
FT   DOMAIN       92    123       COIL 1A.
FT   DOMAIN      124    136       LINKER 1.
FT   DOMAIN      137    232       COIL 1B.
FT   DOMAIN      233    251       LINKER 12.
FT   DOMAIN      252    270       COIL 2A.
FT   DOMAIN      271    279       LINKER 2.
FT   DOMAIN      280    395       COIL 2B.
FT   DOMAIN      396    442       TAIL, SUBDOMAIN A.
FT   DOMAIN      443    548       TAIL, SUBDOMAIN B (ACIDIC).
FT   CARBOHYD     20     20       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   CARBOHYD     26     26       O-LINKED (GLCNAC) (BY SIMILARITY).
FT   SITE        380    390       EPITOPE (RECOGNIZED BY IF-SPECIFIC
FT                                MONOCLONAL ANTIBODY).
FT   UNSURE      322    322       OR K.
SQ   SEQUENCE   548 AA;  61940 MW;  83044813637AC739 CRC64;
```

```
  Query Match             15.4%;  Score 99.5;  DB 1;  Length 548;
  Best Local Similarity   26.2%;  Pred. No. 2.1;
  Matches   34;  Conservative   17;  Mismatches   54;  Indels   25;  Gaps    4;

Qy       5 GRSTY------------RPRPRRY--------VEPPEMIGPMRPEQFSDEVEPATPEEGE 44
           ||| |            |  |  |        :|  | |   : |:  ||  |:  |  |
Db     419 GRSAYGGLQTSSYLMSTRSFPSYYTSHVQEEQIEVEETIEAAKAEEAKDE-PPSEGEAEE 477

Qy      45 PATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEV 104
             ::::  |  | |:|||   :    :  |:|    : |    :|  |:|     ||
Db     478 EGKEKEEAEAEAEAEEEGAQEEEEAAEKEESEEAKEEEGG----EGEQGEETKEAEEEEK 533

Qy     105 KTPEEGEKQS 114
           |    ||:|:
Db     534 KDEGAGEEQA 543
```

```
RESULT  13
GLT5_WHEAT
ID   GLT5_WHEAT     STANDARD;      PRT;   839 AA.
AC   P10388;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutenin, high molecular weight subunit DX5 precursor.
GN   GLU-1D-1D OR GLU-D1-1B.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. CHEYENNE;
RX   MEDLINE=89098419; PubMed=2563152;
RA   Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA   Malpica-Romero J.M.;
RT   "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT   from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT   Cheyenne.";
RL   Nucleic Acids Res. 17:461-462(1989).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RC   STRAIN=CV. CHEYENNE;
RA   Anderson O.D.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC       PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC       VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC   -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC   -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC       GROUP 1 CHROMOSOMES OF WHEAT.
CC   -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC       NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQQPGQ AND
CC       GQQPGQGQQGYYPTS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X12928; CAA31395.1; -.
DR   PIR; S02262; S02262.
DR   InterPro; IPR001419; Glutenin.
DR   PRINTS; PR00210; GLUTENIN.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    839       GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT                                 DX5.
FT   DOMAIN      131    801       REPEATS.
SQ   SEQUENCE   839 AA;  89359 MW;  0F14E1106D552643 CRC64;


  Query Match             15.3%;  Score 99;  DB 1;  Length 839;
  Best Local Similarity   23.5%;  Pred. No. 3.3;
  Matches   31;  Conservative   22;  Mismatches   57;  Indels   22;  Gaps    5;

Qy       4 RGRSTYRP----------RPRRYVEPPE------MIGPMRPEQFS---DEVEPATPEEGE 44
           :|:  | |           :| :: :| :       |::| |     |   |  | :|:
Db     491 QGQPGYYPTSPQQSGQGQQPGQWQQPGQGQPGYYPTSPLQPGQGQPGYDPTSPQQPGQGQ 550

Qy      45 PATQRQDPAAAQEGED-EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEE 103
             | | ||  |:|:       || |      |:    | | :   |  ||:  |   ::
Db     551 QPGQLQQPAQGQQGQQLAQGQQGQQPAQVQQGQQPAQGQQGQQLGQGQQGQQ--PGQGQQ 608

Qy     104 VKTPEEGEKQSQ 115
           : | :|::  |
Db     609 GQQPAQGQQGQQ 620


RESULT  14
GLT4_WHEAT
ID   GLT4_WHEAT     STANDARD;      PRT;   838 AA.
AC   P08489;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutenin, high molecular weight subunit PW212 precursor.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. YAMHILL;
RX   MEDLINE=86093674; PubMed=3001648;
RA   Sugiyama T., Rafalski A., Peterson D., Soll D.G.;
RT   "A wheat HMW glutenin subunit gene reveals a highly repeated
RT   structure.";
RL   Nucleic Acids Res. 13:8729-8737(1985).
CC   -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC       PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC       VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC   -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC   -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC       GROUP 1 CHROMOSOMES OF WHEAT.
CC   -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
CC       NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQQPGQ AND
CC       GQQPGQGQQGYYPTS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X03346; CAA27052.1; -.
DR   PIR; A24107; EEWTHW.
DR   InterPro; IPR001419; Glutenin.
DR   PRINTS; PR00210; GLUTENIN.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    838       GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT                                 PW212.
FT   DOMAIN      130    799       REPEATS.
SQ   SEQUENCE   838 AA;  89174 MW;  71D715B7BDF0722D CRC64;


  Query Match             15.2%;  Score 98;  DB 1;  Length 838;
  Best Local Similarity   22.8%;  Pred. No. 3.9;
  Matches   31;  Conservative   22;  Mismatches   59;  Indels   24;  Gaps    5;

Qy       4 RGRSTYRP----------RPRRYVEPPE------MIGPMRPEQFS---DEVEPATPEEGE 44
           :|:  | |           :| :: :| :       |::| |     |   |  | :|:
Db     496 QGQPGYYPTSPQQSGQGQQPGQWQQPGQGQPGYYPTSPLQPGQGQPGYDPTSPQQPGQGQ 555

Qy      45 PATQRQDPAAAQEGED-EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQE----MDPP 99
             | | ||  |:|:       || |      |:    | | :   |  ||:      |
Db     556 QPGQLQQPAQGQQGQQLAQGQQGQQPAQVQQGQQPAQGQQGQQLGQGQQGQQPGQGQQPA 615

Qy     100 NPEEVKTPEEGEKQSQ 115
           :: : | :|::  |
Db     616 QGQQGQQPGQGQQGQQ 631


RESULT  15
PRPC_HUMAN
ID   PRPC_HUMAN     STANDARD;      PRT;   166 AA.
AC   P02810;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-
DE   3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains:
DE   Peptide P-C].
GN   PRH1 AND PRH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).
RX   MEDLINE=86196106; PubMed=3009472;
RA   Kim H.-S., Maeda N.;
RT   "Structures of two HaeIII-type genes in the human salivary
RT   proline-rich protein multigene family.";
RL   J. Biol. Chem. 261:6712-6718(1986).
RN   [2]
RP   SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).
RX   MEDLINE=85289325; PubMed=2993301;
RA   Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT   "Differential RNA splicing and post-translational cleavages in the
RT   human salivary proline-rich protein gene system.";
RL   J. Biol. Chem. 260:11123-11130(1985).
RN   [3]
RP   SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).
RX   MEDLINE=89061650; PubMed=3196309;
RA   Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,
RA   Madapallimattam G., Schluckebier S.K.;
RT   "The primary structures of six human salivary acidic proline-rich
RT   proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";
RL   Biochem. J. 255:15-21(1988).
RN   [4]
RP   SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).
RX   MEDLINE=88074309; PubMed=3687941;
RA   Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;
RT   "Alleles at the PRH1 locus coding for the human salivary-acidic
RT   proline-rich proteins Pa, Db, and PIF.";
RL   Am. J. Hum. Genet. 41:1035-1047(1987).
RN   [5]
RP   SEQUENCE OF 17-166 (PRP-2).
RX   MEDLINE=86222916; PubMed=3710693;
```

```
RA   Schlesinger D.H., Hay D.I.;
RT   "Complete covalent structure of a proline-rich phosphoprotein, PRP-2,
RT   an inhibitor of calcium phosphate crystal growth from human parotid
RT   saliva.";
RL   Int. J. Pept. Protein Res. 27:373-379(1986).
RN   [6]
RP   SEQUENCE OF 17-166 (PROTEIN C).
RX   MEDLINE=80204368; PubMed=7380845;
RA   Wong R.S.C., Bennick A.;
RT   "The primary structure of a salivary calcium-binding proline-rich
RT   phosphoprotein (protein C), a possible precursor of a related
RT   salivary protein A.";
RL   J. Biol. Chem. 255:5943-5948(1980).
RN   [7]
RP   SEQUENCE OF 17-46 (PROTEIN C).
RX   MEDLINE=81191179; PubMed=7228490;
RA   Schlesinger D.H., Hay D.I.;
RT   "Primary structure of the active tryptic fragments of human and
RT   monkey salivary anionic proline-rich proteins.";
RL   Int. J. Pept. Protein Res. 17:34-41(1981).
RN   [8]
RP   SEQUENCE OF 17-122 (PROTEIN A).
RX   MEDLINE=79173237; PubMed=438215;
RA   Wong R.S.C., Hofmann T., Bennick A.;
RT   "The complete primary structure of a proline-rich phosphoprotein from
RT   human saliva.";
RL   J. Biol. Chem. 254:4800-4808(1979).
RN   [9]
RP   SEQUENCE OF 17-122 (PROTEIN A).
RA   Schlesinger D.H., Hay D.I.;
RT   "Complete primary structure of a proline-rich phosphoprotein (PRP-4),
RT   a potent inhibitor of calcium phosphate precipitation in human parotid
RT   saliva.";
RL   (In) Gross E., Meienhofer J. (eds.);
RL   Peptides: structure and biological function (Proceedings of the 6th
RL   American peptide symposium), pp.133-136, Pierce Chemical Co.,
RL   Rockford Il. (1979).
RN   [10]
RP   SEQUENCE OF 123-166 (PEPTIDE P-C).
RX   MEDLINE=80227634; PubMed=7390979;
RA   Isemura S., Saitoh E., Sanada K.;
RT   "The amino acid sequence of a salivary proline-rich peptide, P-C, and
RT   its relation to a salivary proline-rich phosphoprotein, protein C.";
RL   J. Biochem. 87:1071-1077(1980).
RN   [11]
RP   VARIANT PRH2-3 LYS-163.
RA   Azen E.A.;
RT   "A frequent mutation in the acidic proline-rich protein gene, PRH2,
RT   causing a Q147K change closely adjacent to the bacterial binding
RT   domain of the cognate salivary PRP (Pr1') in Afro-Americans.";
RL   Hum. Mutat. 12:72-72(1998).
CC   -!- FUNCTION: PRP'S ACT AS HIGHLY POTENT INHIBITORS OF CRYSTAL GROWTH
CC       OF CALCIUM PHOSPHATES. THEY PROVIDE A PROTECTIVE AND REPARATIVE
CC       ENVIRONMENT FOR DENTAL ENAMEL WHICH IS IMPORTANT FOR THE INTEGRITY
CC       OF THE TEETH.
CC   -!- PTM: PRP-4 IS DERIVED FROM PRP-2 BY POST-TRANSLATIONAL CLEAVAGE;
CC       PRP-3 (PROTEIN A) IS DERIVED FROM PRP-1 BY POST-TRANSLATIONAL
CC       CLEAVAGE. PIF-F IS DERIVED FROM PIF-S BY POST-TRANSLATIONAL
CC       CLEAVAGE.
CC   -!- POLYMORPHISM: ALLELE PRH1-4 IS ALSO KNOWN AS PRH1 DA ALLELE;
CC       ALLELE PRH2-1 IS ALSO KNOWN AS PR1 OR PROTEIN C; ALLELE PRH2-3 IS
CC       ALSO KNOWN AS PR1'.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K03202; AAA60183.1; -.
DR   EMBL; K03203; AAA60184.1; -.
DR   EMBL; M13057; AAA98807.1; -.
DR   EMBL; M13058; AAA98808.1; -.
DR   PIR; A03290; PIHUSC.
DR   PIR; JP0106; JP0106.
DR   MIM; 168730; -.
DR   MIM; 168790; -.
DR   MIM; 168710; -.
KW   Repeat; Saliva; Signal; Parotid gland; Phosphorylation; Polymorphism.
FT   SIGNAL        1     16
FT   CHAIN        17    166       SALIVARY ACIDIC PROLINE-RICH
FT                                 PHOSPHOPROTEIN 1/2.
FT   CHAIN        17    122       SALIVARY ACIDIC PROLINE-RICH
FT                                 PHOSPHOPROTEIN 3/4.
FT   CHAIN       123    166       PEPTIDE P-C.
FT   DOMAIN       17     46       INHIBIT HYDROXYAPATITE FORMATION, BIND
FT                                 TO HYDROXYAPATITE & CALCIUM.
FT   MOD_RES      17     17       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES      24     24       PHOSPHORYLATION.
FT   MOD_RES      38     38       PHOSPHORYLATION.
FT   VARIANT      20     20       D -> N (IN ALLELE PRH1-4).
FT                                 /FTId=VAR_005563.
FT   VARIANT      66     66       D -> N (IN ALLELE PRH2-1).
FT                                 /FTId=VAR_005564.
FT   VARIANT     163    163       Q -> K (IN ALLELE PRH2-3).
FT                                 /FTId=VAR_005565.
FT   CONFLICT     41     41       F -> P (IN REF. 10).
SQ   SEQUENCE   166 AA;  17017 MW;  A7DF62BF94E3C3EF CRC64;


  Query Match             15.1%;  Score 97.5;  DB 1;  Length 166;
  Best Local Similarity   29.3%;  Pred. No. 1;
  Matches   29;  Conservative    7;  Mismatches   38;  Indels   25;  Gaps    4;

Qy      29 EQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAG---------QGP-----KPEAH 74
           ||| ||      |  |    |:  |:|    :|:|   |         |||     ||:
Db      39 EQFIDEERQGPPLGG----QQSQPSAGDGNQDDGPQQGPPQQGGQQQQGPPPPQGKPQGP 94

Qy      75 SQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           |: |||         | |:   ||       |  |  |
Db      95 PQQGGHPPP-------PQGRPQGPPQQGGHPRPPRGRPQ 126


Search completed: July  1, 2002, 06:35:58
Job time: 199 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:32:13 ; Search time 21.85 Seconds
(without alignments)
918.417 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	637	98.6	116	4	Q9UEU5	Q9ueu5 homo sapien
2	266	41.2	146	4	Q9BSS7	Q9bss7 homo sapien
3	229.5	35.5	111	4	Q96GT9	Q96gt9 homo sapien
4	119	18.4	1300	12	O36421	O36421 alcelaphine
5	117.5	18.2	128	5	O18563	O18563 strongyloid
6	116.5	18.0	956	10	Q9LJ64	Q9lj64 arabidopsis
7	115.5	17.9	144	5	O18564	O18564 strongyloid
8	112	17.3	504	11	Q91X93	Q91x93 mus musculu
9	111	17.2	1054	11	Q9JM99	Q9jm99 mus musculu
10	110	17.0	139	11	Q62882	Q62882 rattus norv
11	110	17.0	221	11	Q62881	Q62881 rattus norv
12	107.5	16.6	110	4	Q96GU1	Q96gu1 homo sapien
13	106	16.4	301	11	Q62105	Q62105 mus musculu
14	105.5	16.3	224	12	O73450	O73450 human papil
15	105.5	16.3	260	11	Q64306	Q64306 mus musculu
16	105.5	16.3	910	11	Q9JLE8	Q9jle8 mus musculu
17	105	16.3	737	10	Q94IL4	Q94il4 secale cere
18	104.5	16.2	912	11	Q9JLE7	Q9jle7 mus musculu
19	104.5	16.2	1343	12	Q06635	Q06635 bovine herp
20	104	16.1	396	5	Q04151	Q04151 toxoplasma
21	104	16.1	926	3	O13305	O13305 pneumocysti
22	102.5	15.9	815	10	Q41553	Q41553 triticum ae
23	102.5	15.9	830	10	Q03872	Q03872 triticum ae
24	102.5	15.9	999	11	Q9JKR6	Q9jkr6 mus musculu
25	102	15.8	480	5	Q27033	Q27033 theileria p
26	102	15.8	614	5	Q94674	Q94674 plasmodium
27	102	15.8	652	11	Q64139	Q64139 mus sp. cbp
28	102	15.8	1560	5	Q26644	Q26644 strongyloce
29	102	15.8	2047	4	O15019	O15019 homo sapien
30	101.5	15.7	300	11	Q61888	Q61888 mus musculu
31	101.5	15.7	543	5	Q27043	Q27043 theileria p
32	101	15.6	897	3	Q96VJ1	Q96vj1 pneumocysti
33	100.5	15.6	847	10	Q9XIB6	Q9xib6 arabidopsis
34	100.5	15.6	1927	2	Q54875	Q54875 streptococc
35	100	15.5	1881	16	Q9L7Q2	Q917q2 streptococc
36	99.5	15.4	309	4	Q04118	Q04118 homo sapien
37	99.5	15.4	543	10	O22120	O22120 glycine max
38	99.5	15.4	605	10	Q94LX2	Q941x2 glycine max
39	99.5	15.4	623	5	Q9W4A2	Q9w4a2 drosophila
40	99.5	15.4	1110	13	Q91255	Q91255 petromyzon
41	99.5	15.4	1130	11	Q91WD8	Q91wd8 mus musculu
42	99	15.3	667	11	O35745	O35745 mus musculu
43	99	15.3	1021	4	O15451	O15451 homo sapien
44	99	15.3	1251	4	O15450	O15450 homo sapien
45	98.5	15.2	215	12	O73449	O73449 human papil

ALIGNMENTS

RESULT 1
Q9UEU5
ID Q9UEU5 PRELIMINARY; PRT; 116 AA.
AC Q9UEU5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GAGE-8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99323388; PubMed=10397259;
RA De Backer O., Arden K.C., Boretti M., Vantomme V., De Smet C.,
RA Czekay S., Viars C.S., De Plaen E., Brasseur F., Chomez P.,
RA Van den Eynde B., Boon T., van der Bruggen P.;
RT "Characterization of the GAGE genes that are expressed in various
RT human cancers and in normal testis.";
RL Cancer Res. 59:3157-3165(1999).
DR EMBL; AF055473; AAC33676.1; -.
SQ SEQUENCE 116 AA; 12764 MW; DD3052939E66F19A CRC64;

Query Match 98.6%; Score 637; DB 4; Length 116;
Best Local Similarity 99.1%; Pred. No. 5.1e-50;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      61 EGASAGQGPKPEADSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
```

RESULT 2

```
Q9BSS7
ID   Q9BSS7      PRELIMINARY;      PRT;   146 AA.
AC   Q9BSS7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   G ANTIGEN, FAMILY B, 1 (PROSTATE ASSOCIATED).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA.;
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC004861; AAH04861.1; -.
SQ   SEQUENCE   146 AA;  16150 MW;  E6C7BA94D898DCB2 CRC64;


  Query Match             41.2%;  Score 266;  DB 4;  Length 146;
  Best Local Similarity   42.0%;  Pred. No. 1.1e-16;
  Matches   66;  Conservative    7;  Mismatches   30;  Indels   54;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           | :  |  || ||  |||           |: |||     |:| |  || ||   |:| ||
Db       1 MGFLRRLIYRRRPMIYVE--------SSEESSDE----QPDEVESPTQSQDSTPAEERED 48

Qy      61 EGASAGQ----------------------------------------GPKPEAHSQEQ 78
           |||||                                          ||:||| ||||
Db      49 EGASAAQGQEPEADSQELVQPKTGCELGDGPDTKRVCLRNEEQMKLPAEGPEPEADSQEQ 108

Qy      79 GHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
           ||:||||  |||| ||:  ||||||||||| | |||
Db     109 VHPKTGCERGDGPDVQELGLPNPEEVKTPEEDEGQSQ 145


RESULT   3
Q96GT9
ID   Q96GT9      PRELIMINARY;      PRT;   111 AA.
AC   Q96GT9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO G ANTIGEN 8.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA, AND CHORIOCARCINOMA;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009232; AAH09232.1; -.
SQ   SEQUENCE   111 AA;  12354 MW;  38C93332C5BA0E14 CRC64;


  Query Match             35.5%;  Score 229.5;  DB 4;  Length 111;
  Best Local Similarity   45.8%;  Pred. No. 1.5e-13;
  Matches   54;  Conservative   15;  Mismatches   38;  Indels   11;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQEG 58
           |||||||||||||| ::|||:|| |         :||    |:| :| |: ::|    |:
Db       1 MSWRGRSTYRPRPRRSLQPPELIGAM--------LEPTDEEPKEEKPPTKSRNPTPDQKR 52

Qy      59 E-DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
           | |:||:  | |  ||  ||    :||  || | | :    |  |  | || || :||
Db      53 EDDQGAAEIQVPDLEADLQELCQTKTGDGCEGGTDVKGKILPKAEHFKMPEAGEGKSQ 110


RESULT   4
O36421
ID   O36421      PRELIMINARY;      PRT;  1300 AA.
AC   O36421;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE IMMEDIATE EARLY PROTEIN.
OS   Alcelaphine herpesvirus 1 (wildebeest herpesvirus).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Gammaherpesvirinae; Rhadinovirus.
OX   NCBI_TaxID=35252;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C500;
RX   MEDLINE=97404659; PubMed=9261371;
RA   Ensser A., Pflanz R., Fleckenstein B.;
RT   "Primary structure of the alcelaphine herpesvirus 1 genome.";
RL   J. Virol. 71:6517-6525(1997).
DR   EMBL; AF005370; AAC58118.1; -.
DR   InterPro; IPR000087; Collagen.
SQ   SEQUENCE   1300 AA;  128183 MW;  40F9EFD244F34577 CRC64;


  Query Match             18.4%;  Score 119;  DB 12;  Length 1300;
  Best Local Similarity   35.2%;  Pred. No. 0.015;
  Matches   45;  Conservative    7;  Mismatches   44;  Indels   32;  Gaps    7;

Qy      17 VEPPEMIGPMRPEQFSDEVEPATPE--EGE-------PATQRQDPAAAQEG--------E 59
           :| ||  ||  ||    |  |  ||  |||       |     |   ||||        |
Db     527 LEGPEGEGPEGPEGPEGE-GPEGPEGPEGEGPEGPEGPERDSPDGPGAQEGPEGPEGPEE 585

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDG------PDGQEMDPPN-------PEEVKT 106
           |||    :||: |     :|    | |  :|      |:| | | |:       || |
Db     586 DEGPEGPEGPEGEGPEGPEGEGPEGLEGPEGDEGPEEPEGPEGDSPDGPGAQEVPEGPKG 645

Qy     107 PEEGEKQS 114
           | ||| ||
Db     646 P-EGECQS 652


RESULT   5
O18563
ID   O18563      PRELIMINARY;      PRT;   128 AA.
AC   O18563;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE   WITH STRONGYLOIDIASIS (FRAGMENT).
OS   Strongyloides stercoralis.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Strongyloididae; Strongyloides.
OX   NCBI_TaxID=6248;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ramachandran S., Thompson R.W., Gam A.A., Neva F.A.;
RT   "A set of recombinant clones for immunodiagnosis of
RT   strongyloidiasis.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U90353; AAB65141.1; -.
DR   InterPro; IPR000087; Collagen.
FT   NON_TER       1       1
SQ   SEQUENCE   128 AA;  13497 MW;  B3BDAFA30BFAE725 CRC64;


  Query Match             18.2%;  Score 117.5;  DB 5;  Length 128;
  Best Local Similarity   34.5%;  Pred. No. 0.0018;
  Matches   38;  Conservative   15;  Mismatches   48;  Indels    9;  Gaps    6;

Qy       6 RSTYRPRPRRYVE-PPEMIGPMRPEQFSDEVEPATPEEGE-PATQRQDPAAAQEGEDEGA 63
           |:  :|:|:   : | |  ||  ||   :   |  ||| | ||   :     |  | ||
Db       4 RAIVKPKPKTTTQAPEEPEGPEEPEGPEEPEGPEGPEEPEGPAGPEEPEGPAGPEEPEGP 63

Qy      64 SAGQGP-KPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEK 112
```

```
          :  || :||   : :|    | |  :||:  |  |  ||| : ||| |:
Db     64 A---GPEEPEGPEEPEG--PAGPEEPEGPEEPE-GPEEPEEPEGPEEPEE 107


RESULT   6
Q9LJ64
ID   Q9LJ64      PRELIMINARY;      PRT;   956 AA.
AC   Q9LJ64;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EXTENSIN PROTEIN-LIKE.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP000735; BAB01698.1; -.
DR   InterPro; IPR001226; Flavodoxin.
DR   InterPro; IPR003592; LRR_out.
DR   InterPro; IPR003882; Pistil_extensin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR01218; PSTLEXTENSIN.
DR   SMART; SM00370; LRR; 5.
DR   PROSITE; PS00201; FLAVODOXIN; UNKNOWN_1.
SQ   SEQUENCE   956 AA;  102822 MW;  E34E185E63BB47C3 CRC64;


  Query Match             18.0%;  Score 116.5;  DB 10;  Length 956;
  Best Local Similarity   27.2%;  Pred. No. 0.018;
  Matches   31;  Conservative   22;  Mismatches   40;  Indels   21;  Gaps    5;

Qy      11 PRPRRYVEPPEMIGPM--------RPEQFSDEV-----EPATPEEGEPATQRQDPAAAQE 57
           | |:: : || :  |          :|:| | :      ||: |:| :| : :|:    ::
Db     418 PEPKKEINPPNLEEPSKPKPEESPKPQQPSPKPETPSHEPSNPKEPKPESPKQESPKTEQ 477

Qy      58 GEDEGASAGQ-GPKPEAHSQEQGHPQTGCECEDGP--DGQEMDPPNPEEVKTPE 108
              : :  |  |  || ||   ||  |:      : |  :  : :|| |||   ||
Db     478 PKPKPESPKQESPKQEAPKPEQPKPK-----PESPKQESSKQEPPKPEESPKPE 526


RESULT   7
O18564
ID   O18564      PRELIMINARY;      PRT;   144 AA.
AC   O18564;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGG AND IGE IMMUNOREACTIVE ANTIGEN RECOGNIZED BY SERA FROM PATIENTS
DE   WITH STRONGYLOIDIASIS (FRAGMENT).
OS   Strongyloides stercoralis.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
OC   Panagrolaimoidea; Strongyloididae; Strongyloides.
OX   NCBI_TaxID=6248;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ramachandran S., Thompson R.W., Gam A.A., Neva F.A.;
RT   "A set of recombinant clones for immunodiagnosis of
RT   strongyloidiasis.";
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U90354; AAB65142.1; -.
DR   InterPro; IPR000087; Collagen.
FT   NON_TER       1       1
SQ   SEQUENCE   144 AA;  15331 MW;  894E20F248405592 CRC64;


  Query Match             17.9%;  Score 115.5;  DB 5;  Length 144;
  Best Local Similarity   34.5%;  Pred. No. 0.0031;
  Matches   38;  Conservative   14;  Mismatches   49;  Indels    9;  Gaps    6;

Qy       6 RSTYRPRPRRYVE-PPEMIGPMRPEQFSDEVEPATPEEGE-PATQRQDPAAAQEGEDEGA 63
           |  :|:|:   : | |  ||  ||   :   |  ||| | ||   :     |  | ||
Db      20 RPIVKPKPKTTTQAPEEPEGPEEPEGPEEPEGPEGPEEPEGPAGPEEPEGPAGPEEPEGP 79

Qy      64 SAGQGP-KPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEK 112
           :   || :||   : :|    | |  :|||:  |  |  ||| : ||| |:
Db      80 A---GPEEPEGPEEPEG--PAGPEEPEGPEEPE-GPEEPEEPEGPEEPEE 123


RESULT   8
Q91X93
ID   Q91X93      PRELIMINARY;      PRT;   504 AA.
AC   Q91X93;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO PROLINE-RICH PROTEIN BSTNI SUBFAMILY 2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SALIVARY GLAND;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011176; AAH11176.1; -.
SQ   SEQUENCE   504 AA;  51775 MW;  E4B9A3317FBEABF2 CRC64;


  Query Match             17.3%;  Score 112;  DB 11;  Length 504;
  Best Local Similarity   29.0%;  Pred. No. 0.023;
  Matches   36;  Conservative   13;  Mismatches   57;  Indels   18;  Gaps    6;

Qy       5 GRSTYRPRPRRYVEPPEMIGPM---RPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDE 61
           |   : | |  :  ||:  ||    ||        |  | :| |  :   |   |    :
Db      81 GPPQHPPHPGHHHGPPQQGPQGPPRPGNQQGPPPPGPPPQGSPQQRPPQPGKQQGPPPQ 140

Qy      62 GASAG-------QGPKPEAHSQEQGHPQTGCECEDGP---DGQEMDPPNP--EEVKTPEE 109
           | | |       ||| ||:   | |  || |   : ||    | :  || |  ::   |:
Db     141 GGSQGPPMPVNQQGPPPKGGPQ-QRPPQPG--NQQGPPPQGGPQQRPPQPGNQQGPPPQG 197

Qy     110 GEKQ 113
           | :|
Db     198 GPQQ 201


RESULT   9
Q9JM99
ID   Q9JM99      PRELIMINARY;      PRT;  1054 AA.
AC   Q9JM99;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 116.0 KDA PROTEIN.
GN   PRG4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20573856; PubMed=11124536;
RA   Ikegawa S., Sano M., Koshizuka Y., Nakamura Y.;
RT   "Isolation, characterization and mapping of the mouse and human PRG4
RT   (proteoglycan 4) genes.";
RL   Cytogenet. Cell Genet. 90:291-297(2000).
DR   EMBL; AB034730; BAA92310.1; -.
DR   MGD; MGI:1891344; Prg4.
DR   InterPro; IPR000585; Hemopexin.
DR   InterPro; IPR002965; P_rich_extensn.
DR   InterPro; IPR001212; Somatomedin_B.
DR   Pfam; PF01033; Somatomedin_B; 2.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   PRINTS; PR00022; SOMATOMEDINB.
DR   SMART; SM00120; HX; 2.
DR   SMART; SM00201; SO; 2.
DR   PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.
DR   PROSITE; PS00524; SOMATOMEDIN_B; 2.
SQ   SEQUENCE   1054 AA;  115992 MW;  4FC64BFA42283235 CRC64;


  Query Match             17.2%;  Score 111;  DB 11;  Length 1054;
  Best Local Similarity   30.3%;  Pred. No. 0.061;
  Matches   36;  Conservative   17;  Mismatches   42;  Indels   24;  Gaps    6;

Qy       7 STYRPRPRRYVEP-------PEMIGPMRPEQFS-DEVEPATPEEGEPATQRQDPAAAQEG 58
           :| :| |    ||       ||   |  ||  :  | || ||:|  | |:: :|   :|
Db     362 TTKKPEPTTRKEPEPTTPKEPEPTTPKEPEPTTPKEPEPTTPKEPPPTTKKPEPTTPKE- 420

Qy      59 EDEGASAGQGP----KPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVK--TPEEGE 111
                 ||    :||  : ::  | |  | |     :| :|  |:| :  ||:| |
Db     421 --------PGPTTPKEPEPTTTKEPEPTTTKEPE-STTRKEPEPTTPKEPEPTTPKEPE 470


RESULT  10
Q62882
ID   Q62882      PRELIMINARY;      PRT;   139 AA.
AC   Q62882;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 15.5 KDA PROTEIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=SKELETAL MUSCLE;
RX   MEDLINE=96221285; PubMed=8634331;
RA   Geertman R., McMahon A., Sabban E.L.;
RT   "Cloning and characterization of cDNAs for novel proteins with
RT   glutamic acid-proline dipeptide tandem repeats.";
RL   Biochim. Biophys. Acta, Gene Struct. Expr. 1306:147-152(1996).
DR   EMBL; U40628; AAB05668.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   139 AA;  15459 MW;  60573DF693989709 CRC64;


  Query Match             17.0%;  Score 110;  DB 11;  Length 139;
  Best Local Similarity   31.8%;  Pred. No. 0.0092;
  Matches   34;  Conservative   15;  Mismatches   36;  Indels   22;  Gaps    5;

Qy      15 RYVEPPEMIGPMRPEQFSDEVEPATPEEGE---PATQRQDPAAAQEGEDEGASAGQGPKP 71
           |  |  |: ||    : |: |:| |||| |    ||:  :|   || | |       |:|
Db      47 RASEEEEIGGP----EDSEAVQPRTPEEPELEAEATKGDEPDLEQEMEPE-------PEP 95

Qy      72 EAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTP--EEGEKQSQC 116
           |   : :  |:   | |  |:      | ||  : |  :||::   |
Db      96 EVEPEPEPEPEPEPEPEPE------PEPEPEREPDFQEGDESEGC 136


RESULT  11
Q62881
ID   Q62881      PRELIMINARY;      PRT;   221 AA.
AC   Q62881;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 24.6 KDA PROTEIN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=PHEOCHROMACYTOMA, AND BRAIN;
RX   MEDLINE=96221285; PubMed=8634331;
RA   Geertman R., McMahon A., Sabban E.L.;
RT   "Cloning and characterization of cDNAs for novel proteins with
RT   glutamic acid-proline dipeptide tandem repeats.";
RL   Biochim. Biophys. Acta, Gene Struct. Expr. 1306:147-152(1996).
DR   EMBL; U40627; AAB05667.1; -.
DR   InterPro; IPR001315; CARD.
DR   SMART; SM00114; CARD; 1.
DR   PROSITE; PS50209; CARD; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   221 AA;  24576 MW;  A7661C9040B2CD4D CRC64;


  Query Match             17.0%;  Score 110;  DB 11;  Length 221;
  Best Local Similarity   31.8%;  Pred. No. 0.015;
  Matches   34;  Conservative   15;  Mismatches   36;  Indels   22;  Gaps    5;

Qy      15 RYVEPPEMIGPMRPEQFSDEVEPATPEEGE---PATQRQDPAAAQEGEDEGASAGQGPKP 71
           |  |  |: ||    : |: |:| |||| |    ||:  :|   || | |       |:|
Db     129 RASEEEEIGGP----EDSEAVQPRTPEEPELEAEATKGDEPDLEQEMEPE-------PEP 177

Qy      72 EAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTP--EEGEKQSQC 116
           |   : :  |:   | |  |:      | ||  : |  :||::   |
Db     178 EVEPEPEPEPEPEPEPEPE------PEPEPEREPDFQEGDESEGC 218


RESULT  12
Q96GU1
ID   Q96GU1      PRELIMINARY;      PRT;   110 AA.
AC   Q96GU1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   UNKNOWN (PROTEIN FOR MGC:16481).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Strausberg R.;
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC009230; AAH09230.1; -.
SQ   SEQUENCE   110 AA;  11777 MW;  6D6AF4563135BB6C CRC64;


  Query Match             16.6%;  Score 107.5;  DB 4;  Length 110;
  Best Local Similarity   32.7%;  Pred. No. 0.012;
  Matches   36;  Conservative   16;  Mismatches   51;  Indels    7;  Gaps    2;

Qy       4 RGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGA 63
           | :|: |   :   :|   :||:  :| ::|      :| || |  |  | : | ::|||
Db       7 RSQSSERGNDQESSQP---VGPVIVQQPTEEKR----QEEEPPTDNQGIAPSGEIKNEGA 59

Qy      64 SAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
           | ||   ||  ||    :      |||| :|   |  :  |  | || |
Db      60 PAVQGTDVEAFQQELALLKIEDAPGDGPDVREGTLPTFDPTKVLEAGEGQ 109
```

```
RESULT  13
Q62105
ID   Q62105      PRELIMINARY;      PRT;   301 AA.
AC   Q62105;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   31-KDA PROLINE-RICH SALIVARY PROTEIN OF CLONE PUMP125.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86033799; PubMed=3840480;
RA   Clements S., Mehansho H., Carlson D.M.;
RT   "Novel multigene families encoding highly repetitive peptide
RT   sequences: Sequence analyses of rat and mouse proline-rich protein
RT   cDNAs.";
RL   J. Biol. Chem. 260:13471-13477(1985).
DR   EMBL; M11897; AAA40001.1; -.
SQ   SEQUENCE   301 AA;  31253 MW;  7F7D99A3F64C6E91 CRC64;


  Query Match             16.4%;  Score 106;  DB 11;  Length 301;
  Best Local Similarity   29.3%;  Pred. No. 0.047;
  Matches   39;  Conservative   10;  Mismatches   54;  Indels   30;  Gaps    7;

Qy       4 RGRSTYR-PRPRRYVEPPEMIGP-MRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDE 61
           :| |  | |:|     ||   ||  || |  ::  |  |  | |  :   |   |    :
Db     121 QGSSQQRPPQPGNQQGPPPQGGPQQRPPQPGNQQGP--PPPGGPQQRPPQPGGNQGPPPQ 178

Qy      62 GASAG-------QGPKPEAHSQ--------EQGHPQTGCECEDGPDG------QEMDPPN 100
             |   |       ||| |:   |        :|| || |     || |      |:  ||
Db     179 GGPQGPPRPGNQQGPPPQGGPQQRPTQPGNQQGPPQQG-----GPQGPPRPGNQQCPPPQ 233

Qy     101 PEEVKTPEEGEKQ 113
                 |  | :|
Db     234 GGPQGPPRPGNQQ 246


RESULT  14
O73450
ID   O73450      PRELIMINARY;      PRT;   224 AA.
AC   O73450;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   E4 PROTEIN (FRAGMENT).
GN   E4.
OS   Human papillomavirus type 76.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=69985;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98118461; PubMed=9454709;
RA   Delius H., Saegling B., Bergmann K., Shamanin V., de Villiers E.M.;
RT   "The genomes of three of four novel HPV types, defined by differences
RT   of their L1 genes, show high conservation of the E7 gene and the
RT   URR.";
RL   Virology 240:359-365(1998).
DR   EMBL; Y15174; CAA75462.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   224 AA;  23747 MW;  475B9582B7A16846 CRC64;


  Query Match             16.3%;  Score 105.5;  DB 12;  Length 224;
  Best Local Similarity   27.0%;  Pred. No. 0.038;
  Matches   33;  Conservative   14;  Mismatches   48;  Indels   27;  Gaps    5;

Qy      12 RPRRYVEPPEMI--------GPMRPEQFSDEVEPATPEEGEPATQRQD---PAAAQEGED 60
             |||  | :||           || ||    :  :|  |    |   :     |   :|| :
Db      17 RPRSYGDPPTPAPFPPPTTPGPRRPPAPQEPPQPTAPPALAPREHQNSHPKPTPGKEGTE 76

Qy      61 E--------------GASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKT 106
           |              | :|  | |     |||  ::  :    || :: || |||: :
Db      77 EKTLALQPPPGKRSRGDAPGPAPGP-GQPPEQGPSESDPKDPADPD-RDPDPENPEDPED 134

Qy     107 PE 108
           ||
Db     135 PE 136


RESULT  15
Q64306
ID   Q64306      PRELIMINARY;      PRT;   260 AA.
AC   Q64306;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   PROLINE-RICH PROTEIN.
GN   PRPMP5 OR PRP MP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=PAROTID;
RX   MEDLINE=92174915; PubMed=1339347;
RA   Layfield R., Bannister A., Pierce E.J., McDonald C.J.;
RT   "cDNA clones for mouse parotid proline-rich proteins. mRNA regulation
RT   by isoprenaline and the nucleotide sequence of proline-rich protein
RT   cDNA MP5.";
RL   Eur. J. Biochem. 204:591-597(1992).
DR   EMBL; X63005; CAA44734.1; -.
DR   EMBL; X63004; CAA44733.1; -.
DR   MGD; MGI:1927478; Prpmp5.
SQ   SEQUENCE   260 AA;  26644 MW;  708D1468106A3002 CRC64;


  Query Match             16.3%;  Score 105.5;  DB 11;  Length 260;
  Best Local Similarity   29.5%;  Pred. No. 0.045;
  Matches   31;  Conservative   15;  Mismatches   50;  Indels    9;  Gaps    4;

Qy      11 PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPK 70
           |||      ||   ||  | :  ::  |  | :| |  |:: |   | |  :|     ||:
Db     144 PRPGNQQGPPPQGGPQGPSRPGNQQGP--PPQGGP--QQRPP---QPGNQQGPPPQGGPQ 196

Qy      71 --PEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQ 113
             |:  :|:   || | :    |  |:  ||     : |: |  |
Db     197 GGPQPGNQQGPPPQGGPQGPPRPGNQQGPPPQGGPQRPPQPGNHQ 241


Search completed: July  1, 2002, 06:35:33
Job time: 200 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 1, 2002, 06:27:08 ; Search time 23.84 Seconds
(without alignments)
540.460 Million cell updates/sec

Title: US-09-782-745-27
Perfect score: 646
Sequence: 1 MSWRGRSTYRPRPRRYVEPP.....DPPNPEEVKTPEEGEKQSQC 116

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	646	100.0	116	21	AAY83159	GAGE2 polypeptide.
2	638	98.8	116	19	AAW47599	GAGE-2 tumour reje
3	623.5	96.5	117	21	AAY83162	GAGE4 polypeptide.
4	619.5	95.9	117	21	AAY83163	GAGE5 polypeptide.
5	616.5	95.4	117	19	AAW47601	GAGE-4 tumour reje
6	615.5	95.3	117	21	AAY83164	GAGE6 polypeptide.
7	613	94.9	118	21	AAY83160	GAGE1 polypeptide.
8	612.5	94.8	117	19	AAW47602	GAGE-5 tumour reje
9	608.5	94.2	117	19	AAW47603	GAGE-6 tumour reje
10	598	92.6	138	19	AAW47598	GAGE-1 tumour reje
11	596.5	92.3	118	21	AAY83161	GAGE3 polypeptide.
12	589.5	91.3	118	19	AAW47600	GAGE-3 tumour reje
13	400	61.9	76	21	AAG02123	Human secreted pro
14	271.5	42.0	111	22	AAM39588	Human polypeptide
15	268.5	41.6	112	22	ABG05297	Novel human diagno
16	266	41.2	146	21	AAB07749	A human cancer-ass
17	249.5	38.6	106	22	AAM41374	Human polypeptide
18	229.5	35.5	111	22	AAM78785	Human protein SEQ
19	229.5	35.5	111	22	AAB60500	Human cell cycle a
20	224	34.7	117	22	AAM79769	Human protein SEQ
21	155.5	24.1	60	22	ABG05299	Novel human diagno
22	141	21.8	115	22	AAE08583	Human NOV4 protein
23	132	20.4	249	22	ABG27048	Novel human diagno
24	120	18.6	102	21	AAY83158	PAGE-4 polypeptide
25	120	18.6	102	21	AAY52998	Human secreted pro
26	113.5	17.6	87	21	AAY83167	PAGE1 polypeptide.
27	109.5	17.0	111	22	AAE08581	Human NOV2 protein
28	108.5	16.8	79	21	AAY83169	PAGE3 polypeptide.
29	104.5	16.2	507	14	AAR37682	gIV from BHV-1 str
30	104	16.1	89	20	AAY12447	Human 5' EST secre
31	103.5	16.0	281	21	AAY90256	Streptococcus equi
32	101	15.6	905	18	AAW31186	Human p160 polypep
33	101	15.6	1135	18	AAW31185	Human p160 polypep
34	100.5	15.6	605	19	AAW62838	Glycine max antimi
35	100	15.5	258	19	AAW55107	Streptococcus pneu
36	100	15.5	392	12	AAR12345	Toxoplasma gondii
37	100	15.5	565	19	AAW61247	Streptococcus pneu
38	100	15.5	1881	21	AAY44506	Streptococcus pneu
39	99.5	15.4	623	22	ABB71084	Drosophila melanog
40	99	15.3	667	19	AAW48760	BOP1 protein. Mus
41	99	15.3	1468	22	ABB62991	Drosophila melanog
42	98.5	15.2	543	22	AAM93466	Human polypeptide,
43	98	15.2	428	12	AAR12352	Toxoplasma gondii
44	98	15.2	1175	22	ABG07962	Novel human diagno
45	97.5	15.1	242	22	ABG00453	Novel human diagno

ALIGNMENTS

RESULT 1
AAY83159
ID AAY83159 standard; Protein; 116 AA.
XX
AC AAY83159;
XX
DT 24-JUL-2000 (first entry)
XX
DE GAGE2 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.
XX
PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues

```
XX
PS   Disclosure; Figure 1a; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. The PAGE
CC   polypeptide shares sequence similarity with the GAGE and MAGE family
CC   of proteins.
XX
SQ   Sequence   116 AA;


  Query Match             100.0%;  Score 646;  DB 21;  Length 116;
  Best Local Similarity   100.0%;  Pred. No. 8.3e-53;
  Matches  116;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 egasagqgpkpeahsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 116


RESULT   2
AAW47599
ID   AAW47599 standard; Protein; 116 AA.
XX
AC   AAW47599;
XX
DT   30-JUL-1998  (first entry)
XX
DE   GAGE-2 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 38
FT                   /note= "Ala encoded by GAG"
FT   Misc-difference 39
FT                   /note= "Thr encoded by CCT"
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   N-PSDB; AAV18717.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 5; 60pp; English.
XX
CC   The present sequence represents a GAGE-2 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein
CC   is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC   The major difference between these proteins and GAGE-1 is the absence
CC   of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC   TRAP sequence. The rest of the sequences show mismatches at various
CC   position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence   116 AA;


  Query Match             98.8%;  Score 638;  DB 19;  Length 116;
  Best Local Similarity   99.1%;  Pred. No. 4.6e-52;
  Matches  115;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      61 egasagqgpkpeaesqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 116


RESULT   3
AAY83162
ID   AAY83162 standard; Protein; 117 AA.
XX
AC   AAY83162;
XX
DT   24-JUL-2000  (first entry)
XX
DE   GAGE4 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
```

```
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1a; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. The PAGE
CC   polypeptide shares sequence similarity with the GAGE and MAGE family
CC   of proteins.
XX
SQ   Sequence   117 AA;


  Query Match             96.5%;  Score 623.5;  DB 21;  Length 117;
  Best Local Similarity   97.4%;  Pred. No. 1e-50;
  Matches  114;  Conservative    1;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 MSWRGRST-YRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||| |||||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyyrprprryvqppemigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      61 degasagggpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 117


RESULT   4
AAY83163
ID   AAY83163 standard; Protein; 117 AA.
XX
AC   AAY83163;
XX
DT   24-JUL-2000  (first entry)
XX
DE   GAGE5 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1a; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. The PAGE
CC   polypeptide shares sequence similarity with the GAGE and MAGE family
CC   of proteins.
XX
SQ   Sequence   117 AA;


  Query Match             95.9%;  Score 619.5;  DB 21;  Length 117;
  Best Local Similarity   96.6%;  Pred. No. 2.4e-50;
  Matches  113;  Conservative    2;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       1 MSWRGRST-YRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||| |||||||||:|||:|||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyyrprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      61 degasagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 117


RESULT   5
AAW47601
ID   AAW47601 standard; Protein; 117 AA.
XX
AC   AAW47601;
XX
DT   30-JUL-1998  (first entry)
XX
DE   GAGE-4 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 39
FT                   /note= "Ala encoded by GAG"
FT   Misc-difference 40
FT                   /note= "Thr encoded by CCT"
XX
```

PN WO9749417-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-US10850.
XX
PR 24-JUN-1996; 96US-0669161.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B;
XX
DR WPI; 1998-076905/07.
DR N-PSDB; AAV18719.
XX
PT Isolated nucleic acid encoding GAGE tumour rejection antigen
PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT diagnose melanomas
XX
PS Example 13; Fig 5; 60pp; English.
XX
CC The present sequence represents a GAGE-4 tumour tumour rejection antigen
CC precursor (TRAP). The protein is expressed in a number of tumours. In
CC contrast the only normal tissue which expresses GAGE TRAP protein
CC is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC The major difference between these proteins and GAGE-1 is the absence
CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC TRAP sequence. The rest of the sequences show mismatches at various
CC position, with the exception of GAGE-3 whose 5' end is totally
CC different from the other GAGE cDNAs for the first 112 bases. This
CC region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC The antigens can be used to diagnose melanomas, characterised by
CC expression of a TRAP or presentation of a tumour rejection antigen.
CC Antigens shed into blood or urine can be observed and then used to
CC confirm a diagnosis of melanoma using cytolytic T cell clone
CC proliferation methodologies. Other uses for the processed peptides,
CC include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ Sequence 117 AA;

Query Match 95.4%; Score 616.5; DB 19; Length 117;
Best Local Similarity 96.6%; Pred. No. 4.6e-50;
Matches 113; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

```
Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  |||||||:|||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyywprprryvqppemigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      61 degasagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 117
```

RESULT 6
AAY83164
ID AAY83164 standard; Protein; 117 AA.
XX
AC AAY83164;
XX
DT 24-JUL-2000 (first entry)
XX
DE GAGE6 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.
XX
PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues
XX
PS Disclosure; Figure 1a; 63pp; English.
XX
CC PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. The PAGE
CC polypeptide shares sequence similarity with the GAGE and MAGE family
CC of proteins.
XX
SQ Sequence 117 AA;

Query Match 95.3%; Score 615.5; DB 21; Length 117;
Best Local Similarity 95.7%; Pred. No. 5.7e-50;
Matches 112; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

```
Qy       1 MSWRGRST-YRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           ||||||||  |||||||||:|||:||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyyyrprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| ||||||||||||||||||:||||||||||||||||||||||||
Db      61 degasagqgpkpeadsqeqghpqtgcecedgpdgqevdppnpeevktpeegekqsqc 117
```

RESULT 7
AAY83160
ID AAY83160 standard; Protein; 118 AA.
XX
AC AAY83160;
XX
DT 24-JUL-2000 (first entry)
XX
DE GAGE1 polypeptide.
XX
KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW immunoconjugate.
XX
OS Homo sapiens.

XX
PN WO200012706-A1.
XX
PD 09-MAR-2000.
XX
PF 31-AUG-1999; 99WO-US20046.
XX
PR 01-SEP-1998; 98US-0098993.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Pastan I, Brinkmann U, Vasmatzis G, Lee B;
XX
DR WPI; 2000-237869/20.
XX
PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT lymphocyte response and for raising antibodies which can be used to
PT detect the presence of PAGE-4 in cell samples or body tissues
XX
PS Disclosure; Figure 1a; 63pp; English.
XX
CC PAGE-4 is a gene preferentially expressed in normal male and female
CC reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC and placenta, as well as in prostate cancer, testicular cancer and
CC uterine cancer. This expression pattern makes it a target for
CC diagnosis and for vaccine based therapy of such neoplasms.
CC An isolated PAGE-4 peptide which induces a cytotoxic T
CC lymphocyte response when bound to a major histocompatibility complex
CC (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC immunogenic compositions to raise a cytotoxic T lymphocyte response
CC against cells expressing PAGE-4 including cancer cells of the
CC prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC PAGE-4 peptide fragments can also be used in these compositions.
CC Antibodies against PAGE-4 and its peptide fragments can be used in
CC detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC samples or body tissues. The presence of PAGE-4 in tissues which are
CC not related to reproduction can be indicative of the spread of
CC cancerous reproductive tissue. PAGE-4 can also be used to raise
CC antibodies which are then used as the targeting group of
CC immunoconjugates comprising toxins used in therapeutic applications.
CC This has applications for drug delivery systems. The PAGE
CC polypeptide shares sequence similarity with the GAGE and MAGE family
CC of proteins.
XX
SQ Sequence 118 AA;

Query Match 94.9%; Score 613; DB 21; Length 118;
Best Local Similarity 97.4%; Pred. No. 9.9e-50;
Matches 111; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQS 114
           |||||||||||| |||||||||||||||||||||||||||||||||||| :|
Db      61 egasagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegemrs 114
```

RESULT 8
AAW47602
ID AAW47602 standard; Protein; 117 AA.
XX
AC AAW47602;
XX
DT 30-JUL-1998 (first entry)
XX
DE GAGE-5 tumour rejection antigen precursor.
XX
KW GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW HLA-typing assay.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 39
FT /note= "Ala encoded by GAG"
FT Misc-difference 40
FT /note= "Thr encoded by CCT"
XX
PN WO9749417-A1.
XX
PD 31-DEC-1997.
XX
PF 23-JUN-1997; 97WO-US10850.
XX
PR 24-JUN-1996; 96US-0669161.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Boon-Falleur T, Debacker O, Van Den Eynde B;
XX
DR WPI; 1998-076905/07.
DR N-PSDB; AAV18720.
XX
PT Isolated nucleic acid encoding GAGE tumour rejection antigen
PT precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT diagnose melanomas
XX
PS Example 13; Fig 5; 60pp; English.
XX
CC The present sequence represents a GAGE-5 tumour tumour rejection antigen
CC precursor (TRAP). The protein is expressed in a number of tumours. In
CC contrast the only normal tissue which expresses GAGE TRAP protein
CC is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC The major difference between these proteins and GAGE-1 is the absence
CC of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC TRAP sequence. The rest of the sequences show mismatches at various
CC position, with the exception of GAGE-3 whose 5' end is totally
CC different from the other GAGE cDNAs for the first 112 bases. This
CC region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC The antigens can be used to diagnose melanomas, characterised by
CC expression of a TRAP or presentation of a tumour rejection antigen.
CC Antigens shed into blood or urine can be observed and then used to
CC confirm a diagnosis of melanoma using cytolytic T cell clone
CC proliferation methodologies. Other uses for the processed peptides,
CC include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ Sequence 117 AA;

Query Match 94.8%; Score 612.5; DB 19; Length 117;
Best Local Similarity 95.7%; Pred. No. 1.1e-49;
Matches 112; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

```
Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  |||||||:|||:||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyywprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db      61 degasagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 117
```

RESULT 9
AAW47603
ID AAW47603 standard; Protein; 117 AA.
XX
AC AAW47603;
XX
DT 30-JUL-1998 (first entry)
XX
DE GAGE-6 tumour rejection antigen precursor.
XX

```
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 39
FT                   /note= "Ala encoded by GAG"
FT   Misc-difference 40
FT                   /note= "Thr encoded by CCT"
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   N-PSDB; AAV18721.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 5; 60pp; English.
XX
CC   The present sequence represents a GAGE-6 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein
CC   is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC   The major difference between these proteins and GAGE-1 is the absence
CC   of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC   TRAP sequence. The rest of the sequences show mismatches at various
CC   position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence   117 AA;

  Query Match             94.2%;  Score 608.5;  DB 19;  Length 117;
  Best Local Similarity   94.9%;  Pred. No. 2.6e-49;
  Matches  111;  Conservative    3;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       1 MSWRGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGE 59
           |||||||||  ||||||||:|||:||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyywprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqege 60

Qy      60 DEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||||| |||||||||||||||||||:||||||||||||||||||||||
Db      61 degasagqgpkpeadsqeqghpqtgcecedgpdgqevdppnpeevktpeegekqsqc 117


RESULT  10
AAW47598
ID   AAW47598 standard; Protein; 138 AA.
XX
AC   AAW47598;
XX
DT   30-JUL-1998  (first entry)
XX
DE   GAGE-1 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 39
FT                   /note= "encoded by GAG"
FT   Misc-difference 40
FT                   /note= "encoded by CCT"
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   N-PSDB; AAV05540.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 7; Fig 5; 60pp; English.
XX
CC   The present sequence represents a GAGE-1 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein
CC   is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC   The major difference between these proteins and GAGE-1 is the absence
CC   of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC   TRAP sequence. The rest of the sequences show mismatches at various
CC   position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence   138 AA;

  Query Match             92.6%;  Score 598;  DB 19;  Length 138;
  Best Local Similarity   99.1%;  Pred. No. 2.9e-48;
  Matches  108;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGED 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mswrgrstyrprprryveppemigpmrpeqfsdevepatpeegepatqrqdpaaaqeged 60

Qy      61 EGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEE 109
           |||||||||||||| |||||||||||||||||||||||||||||||||||
Db      61 egasagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpee 109


RESULT  11
AAY83161
ID   AAY83161 standard; Protein; 118 AA.
XX
```

```
AC   AAY83161;
XX
DT   24-JUL-2000  (first entry)
XX
DE   GAGE3 polypeptide.
XX
KW   PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
KW   uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
KW   cytotoxic T lymphocyte; immunce response; antibody; drug delivery;
KW   immunoconjugate.
XX
OS   Homo sapiens.
XX
PN   WO200012706-A1.
XX
PD   09-MAR-2000.
XX
PF   31-AUG-1999;   99WO-US20046.
XX
PR   01-SEP-1998;   98US-0098993.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Pastan I,  Brinkmann U,  Vasmatzis G,  Lee B;
XX
DR   WPI; 2000-237869/20.
XX
PT   Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T
PT   lymphocyte response and for raising antibodies which can be used to
PT   detect the presence of PAGE-4 in cell samples or body tissues
XX
PS   Disclosure; Figure 1a; 63pp; English.
XX
CC   PAGE-4 is a gene preferentially expressed in normal male and female
CC   reproductive tissues e.g. prostate, testis, fallopian tube, uterus
CC   and placenta, as well as in prostate cancer, testicular cancer and
CC   uterine cancer. This expression pattern makes it a target for
CC   diagnosis and for vaccine based therapy of such neoplasms.
CC   An isolated PAGE-4 peptide which induces a cytotoxic T
CC   lymphocyte response when bound to a major histocompatibility complex
CC   (MHC) class I molecule or the isolated PAGE-4 protein can be used in
CC   immunogenic compositions to raise a cytotoxic T lymphocyte response
CC   against cells expressing PAGE-4 including cancer cells of the
CC   prostate, uterus and testis. The nucleic acids encoding PAGE-4 or
CC   PAGE-4 peptide fragments can also be used in these compositions.
CC   Antibodies against PAGE-4 and its peptide fragments can be used in
CC   detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell
CC   samples or body tissues. The presence of PAGE-4 in tissues which are
CC   not related to reproduction can be indicative of the spread of
CC   cancerous reproductive tissue. PAGE-4 can also be used to raise
CC   antibodies which are then used as the targeting group of
CC   immunoconjugates comprising toxins used in therapeutic applications.
CC   This has applications for drug delivery systems. The PAGE
CC   polypeptide shares sequence similarity with the GAGE and MAGE family
CC   of proteins.
XX
SQ   Sequence   118 AA;

  Query Match             92.3%;  Score 596.5;  DB 21;  Length 118;
  Best Local Similarity   95.6%;  Pred. No. 3.4e-48;
  Matches  109;  Conservative    3;  Mismatches    1;  Indels    1;  Gaps    1;

Qy       4 RGRST-YRPRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEG 62
           ||:|| |||||||||:|||:||||||||||||||||||||||||||||||||||||||||
Db       5 rgkstyyrprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqegedeg 64

Qy      63 ASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      65 asagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 118

RESULT  12
```

```
AAW47600
ID   AAW47600 standard; Protein; 118 AA.
XX
AC   AAW47600;
XX
DT   30-JUL-1998  (first entry)
XX
DE   GAGE-3 tumour rejection antigen precursor.
XX
KW   GAGE tumour tumour rejection antigen precursor; TRAP; tumour;
KW   diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
KW   HLA-typing assay.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 40
FT                   /note= "Ala encoded by GAG"
FT   Misc-difference 41
FT                   /note= "Thr encoded by CCT"
XX
PN   WO9749417-A1.
XX
PD   31-DEC-1997.
XX
PF   23-JUN-1997;   97WO-US10850.
XX
PR   24-JUN-1996;   96US-0669161.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Boon-Falleur T,  Debacker O,  Van Den Eynde B;
XX
DR   WPI; 1998-076905/07.
DR   N-PSDB; AAV18718.
XX
PT   Isolated nucleic acid encoding GAGE tumour rejection antigen
PT   precursor - processed by HLA-Cw6 molecules into peptides, useful to
PT   diagnose melanomas
XX
PS   Example 13; Fig 5; 60pp; English.
XX
CC   The present sequence represents a GAGE-3 tumour tumour rejection antigen
CC   precursor (TRAP). The protein is expressed in a number of tumours. In
CC   contrast the only normal tissue which expresses GAGE TRAP protein
CC   is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC   The major difference between these proteins and GAGE-1 is the absence
CC   of a stretch of 143 bases located at position 379 to 521 of the GAGE-1
CC   TRAP sequence. The rest of the sequences show mismatches at various
CC   position, with the exception of GAGE-3 whose 5' end is totally
CC   different from the other GAGE cDNAs for the first 112 bases. This
CC   region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC   The antigens can be used to diagnose melanomas, characterised by
CC   expression of a TRAP or presentation of a tumour rejection antigen.
CC   Antigens shed into blood or urine can be observed and then used to
CC   confirm a diagnosis of melanoma using cytolytic T cell clone
CC   proliferation methodologies. Other uses for the processed peptides,
CC   include HLA-typing assays for, e.g. skin graft or organ transplants.
XX
SQ   Sequence   118 AA;

  Query Match             91.3%;  Score 589.5;  DB 19;  Length 118;
  Best Local Similarity   94.7%;  Pred. No. 1.5e-47;
  Matches  108;  Conservative    3;  Mismatches    2;  Indels    1;  Gaps    1;

Qy       4 RGRSTYR-PRPRRYVEPPEMIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEG 62
           ||:|||  |||||||:|||:||||||||||||||||||||||||||||||||||||||||
Db       5 rgkstyywprprryvqppevigpmrpeqfsdevepatpeegepatqrqdpaaaqegedeg 64

Qy      63 ASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQC 116
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      65 asagqgpkpeadsqeqghpqtgcecedgpdgqemdppnpeevktpeegekqsqc 118.
```

```
RESULT  13
AAG02123
ID   AAG02123 standard; Protein; 76 AA.
XX
AC   AAG02123;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 6204.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC02129.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 6204; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   76 AA;


  Query Match             61.9%;  Score 400;  DB 21;  Length 76;
  Best Local Similarity   97.3%;  Pred. No. 4e-30;
  Matches   73;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      22 MIGPMRPEQFSDEVEPATPEEGEPATQRQDPAAAQEGEDEGASAGQGPKPEAHSQEQGHP 81
           ||||||||||||||||||||||||||| ||||||| |||||||||||||||||||||||||
Db       1 migpmrpeqfsdevepatpeegepatqxqdpaaaqxgedegasagqgpkpeahsqeqghp 60

Qy      82 QTGCECEDGPDGQEM 96
           ||||||||||||||||
Db      61 qtgcecedgpdgqem 75


RESULT  14
AAM39588
ID   AAM39588 standard; Protein; 111 AA.
XX
AC   AAM39588;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Human polypeptide SEQ ID NO 2733.
XX
KW   Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW   peripheral nervous system; neuropathy; central nervous system; CNS;
KW   Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW   amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW   chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW   leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200153312-A1.
XX
PD   26-JUL-2001.
XX
PF   26-DEC-2000; 2000WO-US34263.
XX
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
PR   09-JUL-2000; 2000US-0598042.
PR   19-JUL-2000; 2000US-0620312.
PR   03-AUG-2000; 2000US-0653450.
PR   14-SEP-2000; 2000US-0662191.
PR   19-OCT-2000; 2000US-0693036.
PR   29-NOV-2000; 2000US-0727344.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Asundi V,  Chen R,  Ma Y,  Qian XB,  Ren F,  Wang D;
PI   Wang J,  Wang Z,  Wehrman T,  Xu C,  Xue AJ,  Yang Y,  Zhang J;
PI   Zhao QA,  Zhou P,  Goodrich R,  Drmanac RT;
XX
DR   WPI; 2001-442253/47.
DR   N-PSDB; AAI58744.
XX
PT   Novel nucleic acids and polypeptides, useful for treating disorders
PT   such as central nervous system injuries -
XX
PS   Example 4; SEQ ID NO 2733; 10078pp; English.
XX
CC   The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC   the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC   immunosuppressant and cytostatic activity. The polynucleotides are useful
CC   in gene therapy. A composition containing a polypeptide or polynucleotide
CC   of the invention may be used to treat diseases of the peripheral nervous
CC   system, such as peripheral nervous injuries, peripheral neuropathy and
CC   localised neuropathies and central nervous system diseases, such as
CC   Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC   lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC   utilisation of the activities such as: Immune system suppression,
CC   Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC   assays for receptor activity, arthritis and inflammation, leukaemias and
CC   C.N.S disorders.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification.
XX
SQ   Sequence   111 AA;


  Query Match             42.0%;  Score 271.5;  DB 22;  Length 111;
  Best Local Similarity   50.0%;  Pred. No. 5.4e-18;
  Matches   59;  Conservative   13;  Mismatches   35;  Indels   11;  Gaps    3;

Qy       1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQE- 57
           | |||||||||||||| | |||:||||        :||    |:: || |: :||| ||
Db       1 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpapgqer 52

Qy      58 GEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
              ||:||:  | |  ||  ||    :|| || :||| |    |  |: | || |::| |
```

```
Db     53 eedqgaaetqvpdleadlqelsqsktggecgngpddqgkilpkseqfkmpeggdrqpq 110


RESULT 15
ABG05297
ID   ABG05297 standard; Protein; 112 AA.
XX
AC   ABG05297;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #5288.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS69484.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 35656; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   112 AA;


  Query Match             41.6%;  Score 268.5;  DB 22;  Length 112;
  Best Local Similarity   49.2%;  Pred. No. 1e-17;
  Matches   58;  Conservative   14;  Mismatches   35;  Indels   11;  Gaps    3;

Qy      1 MSWRGRSTYRPRPRRYVEPPEMIGPMRPEQFSDEVEPA--TPEEGEPATQRQDPAAAQE- 57
          | ||||||||||||| | |||:|||||        :||    |:: || |: :|||  ||
Db      2 miwrgrstyrprprrsvpppeligpm--------lepgdeepqqeepptesrdpapgqer 53

Qy     58 GEDEGASAGQGPKPEAHSQEQGHPQTGCECEDGPDGQEMDPPNPEEVKTPEEGEKQSQ 115
          ||:|::  | |  ||  ||    :|| || :||| |    |  |: | || |::| |
Db     54 kedqgsaktqvpdleadlqelsqsktggecgngpddqgkilpkseqfkmpeggdrqpq 111


Search completed: July  1, 2002, 06:32:09
Job time: 301 sec
```